University of Alberta

Artificial Neural Networks Modelling of Streamflow and Water Quality in Ungauged Watersheds: Investigating the Potential of Remote Sensing Information

by

Mohamed H. Nour ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of **Doctor of Philosophy**

in

Environmental Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta

Fall, 2007

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33038-8
Our file Notre référence
ISBN: 978-0-494-33038-8

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

DEDICATION

To my wife,

To my son,

To my parents and grandparents.

ABSTRACT

Most of the currently available models for watershed modelling are limited in practice because of the extensive requirement for landscape data. A class of models that can simulate the response of ungauged watersheds, without being ground-based data collection and time intensive, was developed to provide the necessary information for responsive watershed management practices.

A class of watershed models that are less reliant on ground-based measurements by using remote sensing (RS) information instead was devised. The focus was on formulating streamflow (Q) and total phosphorus (TP) concentrations models, which are only reliant on the currently available meteorological information, as well as public-domain free-of-cost Moderate Resolution Imaging Spectroradiometer (MODIS)-derived RS information. A number of Q and TP models were devised and applied to a number of watersheds (5 to 130 km^2 in basin area.) The thesis presented: (1) the first effort to compare autoregressive moving average with exogenous input (ARMAX) modelling to artificial neural network (ANN) modelling for TP predictions and confirmed that the ANN approach is superior to the ARMAX in modelling time-correlated gapped data; (2) a step-by-step guideline to ANN modelling of time-correlated variables that can account for data hystereses; (3) an ANN modelling algorithm that relies only on low-cost, readily available meteorological data, and careful time series manipulation prior to model building for Q predictions and, thus, is suitable for modelling streamflow in ungauged watersheds; (4) a new remotely-sensed hydrologic similarity measure that provided a successful indicator of basins similarity; (5) a successful modelling algorithm that can rely on a dynamic suite of RS vegetation indices (VIs) for predicting

TP concentrations and; (6) the first attempt to address the impact of watershed subdivision on a water quality parameter using an ANN modelling technique.

The results from this exercise demonstrated the applicability of the ANN modelling approach, and the usefulness of the MODIS-derived VIs in simulating Q and TP dynamics. Such models can potentially serve as valuable tools for watershed-scale forest management.

ACKNOWLEDGEMENTS

First and foremost I am thankful to Allah the Almighty for his grace and mercy. I would like to express my sincere gratitude to my supervisor Dr. D.W. Smith for his patience, guidance, support, and continuous encouragement throughout the course of this study. Dr. Smith devoted his time and effort to make this study a success. His interest in the subject and in my general well being, contributed very significantly to making this research work possible. I also want to offer my appreciation to my co-supervisor Dr. M. Gamal El-Din for his important contributions to my work, and for his continuous encouragement and support throughout.

I am thankful to my committee members Dr. F. Anctil, Dr. C. Mendoza, Dr. T. Gan, Dr. I.D. Buchanan, and Dr. Q. Zhang for their suggestions that made this thesis a better manuscript.

This thesis is part of a bigger project named the Forest Watershed and Riparian Disturbance (FORWARD) study. The FORWARD contributions to this work were invaluable. I am truly grateful to Dr. E.E. Prepas and Dr. G. Putz for the numerous discussions and the constructive feedback they offered during the course of this study. I am thankful to Mr. J. Russell for his encouragements and for teaching me how to make my work of industrial benefit. Without the efforts of Mr. M. Serediak, Mr. S. Pinder, Dr. I. Whitson, and Ms. J. Burke in data collection and preparation, none of the models would have been developed. I really appreciate their help.

The financial support from the following organizations for the FORWARD Project is gratefully acknowledged: the Natural Sciences and Engineering Research Council of

Canada (NSERC) Collaborative Research and Development program, Millar Western Forest Products Ltd., Blue Ridge Lumber Inc. (a Division of West Fraser Timber Company Ltd.), Alberta Newsprint Company (ANC Timber), Vanderwell Contractors (1971) Ltd., the Canada Foundation for Innovation, the Living Legacy Research Program, and the Ontario Innovation Trust.

I was very fortunate to receive a number of scholarships and prizes during the term of this work. I am very thankful to the organizing and the selection committees of the following scholarships: John and Patricia Schlosser Environment Scholarship, Dorothy J. Killam Memorial Graduate Prize, Izaak Walton Killam Memorial Scholarship, FS Chia PhD Scholarship, J. Gordon Kaplan Graduate Student Award, Andrew Stewart Memorial Graduate Prize, Inland Cement Graduate Scholarship in Environmental Studies, and Dr. Donald R. Stanley Graduate Scholarship in Environmental Engineering.

I would like to thank Dr. C. Deutsch and Dr. O. Leuangthong for helping me in the geostatistical component of this thesis. I am also grateful to Dr. Sanchez-Azofeifa for giving me access to the Earth Observation Systems Laboratory at the University of Alberta for conducting the image processing component of this work.

The friends that I have made and had the opportunity to work with have enriched my experience considerably. Mahad Baawain and Tarek Abdelaziz were more than friends throughout this work. Their moral support made it possible for me to finish this work. I would like to truthfully thank them for all the help and support they offered. My friends and colleagues, Afnan Khan, Steve Pickle, and Clark Svrcek, have been a constant source of support and encouragement throughout my research.

Words cannot express my love and gratitude for my wife Amena, my dad Hamdy, my mum Alaa, my sister Rania, and my brother Aly. Without their support, I could have never had the dedicated time necessary to complete my research. I only hope one day to repay their dedication. Although my son Ismail is still young, I hope he can one day understand and forgive me for not spending enough time with him.

TABLE OF CONTENTS

CHAPTER 1. BACKGROUND AND GENERAL INTRODUCTION 1
- 1.1. Background 1
- 1.2. Watershed Modelling 2
  - 1.2.1. Watershed models classification 2
  - 1.2.2. Conceptual Watershed Q and P Models 5
  - 1.2.3. Data-driven Q and P Models 6
- 1.3. Application of Remote Sensing in Watershed Modelling 8
- 1.4. Research Needs 10
- 1.5. Research Objectives 11
- 1.6. Research Contributions 12
- 1.7. Thesis Organization 13
- 1.8. References 16

CHAPTER 2. ARTIFICIAL NEURAL NETWORKS AND TIME SERIES MODELLING OF TP CONCENTRATION IN BOREAL STREAMS: A COMPARATIVE APPROACH 26
- 2.1. Introduction 26
- 2.2. Study Area and Database 29
- 2.3. Methodology 30
- 2.4. ARMAX TP Modelling 31
- 2.5. ANN TP Modelling 38
- 2.6. Discussion and Implications 44
- 2.7. Conclusions and Final Remarks 47
- 2.8. References 49

CHAPTER 3. THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS TO Q AND ΔTP IN SMALL STREAMS ON THE BOREAL PLAIN, WITH EMPHASIS ON THE ROLE OF WETLANDS 66

3.1. Introduction 66

3.2. Time Series Analysis 70

3.3. Artificial Neural Networks 72

3.4. Development of ANN Models 74

3.4.1. Data Pre-processing Phase 75

3.4.2. Model Building Phase 80

3.4.3. Model Evaluation Phase 83

3.5. Results and Discussion 84

3.5.1. Case study 1: The Willow Creek watershed 84

3.5.2. Case study 2: 1A Creek Watershed 86

3.6. Conclusions and Recommendations 87

3.7. References 88

CHAPTER 4. GEOSTATISTICAL MAPPING OF PRECIPITATION: IMPLICATIONS FOR RAIN GAUGE NETWORK DESIGN 102

4.1. Introduction 102

4.2. Study Area and Data Analysis 103

4.3. Variography and Variogram Modelling 104

4.4. Kriging Techniques 106

4.4.1. Simple Kriging (SK) 107

4.4.2. Ordinary Kriging (OK) 108

4.4.3. Multi-Gaussian Kriging (MGK) 109

4.4.4. Log-Normal Kriging (LNK) 110

4.4.5. Kriging with an External Drift (KED) 110

4.5. Inverse Distance Weights (IDW) Interpolation 112

4.6. Evaluation of The Different Interpolation Methods 113

4.7. Sequential Gaussian Simulation (SGS) 114
4.8. Future Rain Gauge Network Design 115
4.9. Conclusions and Recommendations 116
4.10. References 118

CHAPTER 5. TOWARDS A GENERIC ARTIFICIAL NEURAL NETWORK MODEL FOR DYNAMIC PREDICTIONS OF DAILY STREAMFLOW IN UNGAUGED WATERSHEDS: INTRODUCING A NEW MEASURE OF HYDROLOGIC SIMILARITY 129
5.1. Introduction 129
5.2. Research Area and Database 131
5.3. Artificial Neural Networks Model Development 132
5.3.1. Data Pre-processing and Input Determination 133
5.3.2. Model Construction 139
5.3.3. Model Evaluation 143
5.4. Model Application 144
5.5. Application to an Ungauged Watershed 147
5.6. Scaling-up and Regionalization of Models 148
5.7. Conclusions and Recommendations 151
5.8. References 152

CHAPTER 6. DEVELOPING A TRAIN OF ARTIFICIAL NEURAL NETWORKS MODELS FOR MODELLING STREAMFLOW, SUSPENDED SOLIDS, AND PHOSPHORUS 170
6.1. Introduction 170
6.2. Materials and Methods 171
6.3. ANN Model Development 172
6.3.1. Determination of Model Inputs 174
6.3.2. Data Division 175
6.3.3. Determination of Network Architecture 175

6.4. Results and Discussion 176
    6.4.1. ANN Q Model 176
    6.4.2. ANN TSS Model 177
    6.4.3. ANN TP Model 177
6.5. Conclusions and Final Remarks 178
6.6. References 179

CHAPTER 7. ON THE POTENTIAL OF SATELLITE DERIVED VEGETATION INDICES FOR WATERSHED PHOSPHORUS MODELLING: A NEURAL NETWORK APPROACH 188

7.1. Introduction 188
7.2. Remotely Sensed Vegetation Indices 191
7.3. Study Area and Data Acquisition 194
    7.3.1. Study Area and Ground-Based Data Acquisition 194
    7.3.2. Remote Sensing Data Acquisition 195
7.4. Vegetation Dynamics and Linkage to TP Modelling 197
7.5. Methodology 199
7.6. Development of Artificial Neural Networks Models 200
    7.6.1. Data Pre-processing and Input Determination Phase 202
    7.6.2. Model Building Phase 204
    7.6.3. Model Evaluation Phase 205
7.7. Results and Discussion 206
    7.7.1. Modelling Results 206
    7.7.2. Sensitivity Analysis 208
7.8. Model Applications 210
7.9. Conclusions and Recommendations 211
7.10. References 213

CHAPTER 8. EFFECT OF WATERSHED SUBDIVISION ON WATER-PHASE PHOSPHORUS MODELLING: AN ARTIFICIAL NEURAL NETWORK MODELLING APPLICATION 238

8.1. Introduction 238

8.2. Study Area and Input Database 242

8.2.1. Ground-Based Data Acquisition 243

8.2.2. Remotely-Sensed Data Acquisition 243

8.3. Watershed Delineation/Discretization 246

8.4. Methodology 247

8.5. ANN Model Building 249

8.5.1. Data Pre-processing 250

8.5.2. Model Formulation 252

8.5.3. Model Assessment 254

8.6. Results and Discussion 256

8.7. Conclusions and Recommendations 261

8.8. References 263

CHAPTER 9. GENERAL CONCLUSIONS AND RECOMMENDATIONS 284

9.1. THESIS OVERVIEW 284

9.2. Conclusions 286

9.3. Recommendations for Future Work 292

LIST OF TABLES

Table 2-1. Summary statistics of the two modelling approaches. 54

Table 3-1. Summary table for all models' inputs 93

Table 3-2. Summary table showing optimum ANN models' architecture and ANN internal parameters 93

Table 3-3. Statistical measures of models' performance 94

Table 4-1. Summary table of weather stations' locations 120

Table 4-2. Summary of variogram models utilized in this study 121

Table 4-3. Statistical performance of utilized kriging algorithms 121

Table 5-1. Summary table for all models' inputs 157

Table 5-2. Summary table showing optimum ANN models' architecture and ANN internal parameters 157

Table 5-3. Statistical measures of models' performance 158

Table 5-4. Statistical measures of models' performance applied to the Mosquito watershed 158

Table 7-1. A summary table of the vegetation indices explored in this study 221

Table 7-2. Summary table showing optimum ANN models' architecture and ANN internal parameters 222

Table 7-3. Statistical measures of models' performances 223

Table 8-1. Summary table showing optimum ANN models' architecture and ANN internal parameters 271

Table 8-2. Statistical measures of models' performance 272

LIST OF FIGURES

Figure 2-1. Daily stream flow hydrograph and TP concentration profile of Willow Creek watershed from 5 May 2001 through 31 October 2002. 54

Figure 2-2. ACF (a) and PACF (b) of TP concentration time series, dotted lines showing upper and lower 95% confidence boundaries. 55

Figure 2-3. ACF (a) and PACF (b) of first difference TP concentration time series, dotted lines showing upper and lower 95% confidence boundaries. 56

Figure 2-4. ACF (a) and PACF (b) of first difference Q time series, dotted lines showing upper and lower 95% confidence boundaries. 57

Figure 2-5. CCF of flow and TP concentration time series after applying first difference to both series, dotted lines showing upper and lower 95% confidence boundaries. 58

Figure 2-6. Smoothed periodogram ($\hat{f}(\upsilon)$) of Q (a) in ($[m^3/s]^2$ and TP concentration (b) in $[\mu g/L]^2$ time series, L=3. 58

Figure 2-7. Squared coherence of flow and TP concentration time series. Dotted lines depict dominant frequency of both series, L=3. 59

Figure 2-8. Measured and ARMAX predicted TP concentration profiles. 59

Figure 2-9. Scatter plot of measured and ARMAX predicted TP concentration. 60

Figure 2-10. ANN model selected architecture. 61

Figure 2-11. Impact of increasing the number of neurons per hidden layer slabs on model performance. 62

Figure 2-12. Scatter plot of Measured and ANN predicted TP concentration for training data set (a) and testing data set (b). 63

Figure 2-13. Measured and ANN predicted TP concentration profiles. 64

Figure 2-14. Relative importance of ANN model inputs. 64

Figure 2-15. Proposed systematic guidelines for modelling time correlated variables using back-propagation ANN. 65

Figure 3-1. The study area showing the two modeled watersheds 94

Figure 3-2. 1A Creek watershed soil map showing the wetlands locations within the watershed 95

Figure 3-3. CCF of 1A Creek flow and *TP* concentration, dashed line is the 95% confidence boundary 96

Figure 3-4. Power spectrum of 1A Creek flow in $[m^3/s]^2$ (left panel) and *TP* concentration in $[\mu g/l]^2$ (right panel) 96

Figure 3-5. Schematic showing ANN optimum architecture for all four models 97

Figure 3-6. Measured versus ANN predicted flow hydrographs for the Willow Creek watershed 98

Figure 3-7. Measured versus ANN predicted *TP* concentration profile for the Willow Creek watershed 99

Figure 3-8. Measured versus ANN predicted flow hydrographs for the 1A Creek watershed 100

Figure 3-9. Measured versus ANN predicted *TP* concentration profile for the 1A Creek watershed 101

Figure 4-1. (a) 30-m resolution DEM and (b) slope distribution in the study area...... 122

Figure 4-2. (a) histogram of weather station pairs of separation distances, (b) wind-rose type histogram of separation distances, and (c) colour coded directional semivariogram 123

Figure 4-3. Experimental and modelled omni-directional semivariograms for (a) original data space, (b) normal-score transformed data, and (c) log-normal transformed data; solid lines represent modelled semivariograms 124

Figure 4-4. Kriging maps of rainfall in mm obtained by OK for (a) June 28, 2002 and (b) July 8, 2004; arrows showing prevailing wind direction 125

Figure 4-5. Estimation variance obtained by OK for (a) June 28, 2002 and (b) July 8, 2004 126

Figure 4-6. Spatial probability distribution of exceeding (a) 50 mm/d and (b) 5 mm/d 127

Figure 4-7. (a) Kriging variance distribution after adding the four FORWARD weather stations (b) DEM, and (c) proposed new sites for future weather stations (black rectangles); ▲ denoting current FORWARD study weather stations 128

Figure 5-1. Study area showing studied watersheds and utilized weather stations (some weather stations are not shown at this scale) 159

Figure 5-2. Schematic of the inverse distance weighted (IDW) interpolation 160

Figure 5-3. CCF (*Q*, *R*) for (a) 1A, (b) Cassidy, (c) Two Creek, and (d) Willow watersheds. Solid line shows the 95% confidence boundary 161

Figure 5-4. Power spectrum of flow time series in $[m^3/s]^2$ for (a) 1A, (b) Cassidy, (c) Two Creek, and (d) Willow watersheds 162

Figure 5-5. A description of the concept of feeding ANN *Q* models with seasonal variation and *Q*/*R* hystereses loops 163

Figure 5-6. Schematic showing the usage of the testing data set in optimizing model training 164

Figure 5-7. Comparison of measured and modelled flow hydrographs of 1A watershed 164

Figure 5-8. Comparison of measured and modelled flow hydrographs of Cassidy watershed 165

Figure 5-9. Comparison of measured and modelled flow hydrographs of Two Creek watershed 165

Figure 5-10. Comparison of measured and modelled flow hydrographs of Willow watershed 166

Figure 5-11. Importance of model inputs for streamflow predictions 167

Figure 5-12. Comparison of measured and modelled flow hydrographs of Cassidy watershed when 1A ANN model was used for Cassidy flow prediction 168

Figure 5-13. A simple linear regression between models' "goodness-of-fit" measures and SWMIR_SI similarity index 168

Figure 5-14. Comparison of measured and modelled flow hydrographs of Mosquito watershed when Willow ANN model (class 1) was used for Mosquito flow predictions 169

Figure 6-1. A schematic showing data division using Kohonen neural networks 182

Figure 6-2. Optimum ANN models' architectures and model inputs 183

Figure 6-3. (a) Top panel; scatter plot of measured and predicted flow values and (b) bottom panel; measured vs. modelled flow hydrographs 184

Figure 6-4. (a) Top panel; scatter plot of measured and predicted TSS values and (b) bottom panel; measured vs. modelled TSS profiles 185

Figure 6-5. (a) Top panel; scatter plot of measured and predicted TP values and (b) bottom panel; measured vs. modelled TP profiles 186

Figure 6-6. A proposed framework for utilizing the developed models in water resources management.. 187

Figure 7-1. Study Area: Willow Creek and Two-Creek watersheds, Alberta, Canada 224

Figure 7-2. Comparison of EVI, NDVI, and NDWI time profiles for The Willow Creek watershed 225

Figure 7-3. Comparison of GFVI and GFVIm time profiles for the Willow Creek watershed 226

Figure 7-4 . Comparison of SRVI and SRVIm time profiles for the Willow Creek watershed 227

Figure 7-5. Power spectrum of TP in $[\mu g/l]^2$ for the Willow watershed 228

Figure 7-6. Trigonometric quadrants describing the concept of feeding ANN TP models with seasonal variation and TP/Q hystereses loops 228

Figure 7-7. A schematic showing ANN optimum architecture for all six models 229

Figure 7-8. Sensitivity of TP predictions to changes in EVI, Rsq = R^2 230

Figure 7-9. Sensitivity of TP predictions to changes in NDVI, Rsq = R^2 231

Figure 7-10. Sensitivity of TP predictions to changes in GFVI, Rsq = R^2 232

Figure 7-11. Sensitivity of TP predictions to changes in GFVIm, Rsq = R^2 233

Figure 7-12. Sensitivity of TP predictions to changes in SRVI, Rsq = R^2 234

Figure 7-13. Sensitivity of TP predictions to changes in SRVIm, Rsq = R^2 235

Figure 7-14. Sensitivity of TP predictions to changes in NDWI, Rsq = R^2 236

Figure 7-15. Measured vs. ANN predicted TP concentration profiles for the ANNTP(EVI) model applied to the Two-Creek watershed 237

Figure 8-1. A schematic showing the studied watershed 273

Figure 8-2. 30 m x 30 m resolution filled DEM of the FORWARD study area 274

Figure 8-3. The FORWARD study area generated flow direction map using the D8 method 275

Figure 8-4. Generated flow accumulation map for the FORWARD study area 276

Figure 8-5. Generated watersheds for the FORWARD study area using a 250 ha TDA 277

Figure 8-6. The Willow watershed subdivisions used in this study 278

Figure 8-7. Trigonometric quadrants describing the concept of feeding ANN TP models with seasonal variation and TP/Q hystereses loops 279

Figure 8-8. ANN architecture used in this study 280

Figure 8-9. Impact of watershed discretizations on models' performance.................. 281

Figure 8-10. TP concentration profiles for all four models applied to the cross-validation data set 281

Figure 8-11. Predictions of TP concentration values > 50 µg/L in the cross-validation data set, SS used as the testing data set (top panel) and CVS used as the testing data set (bottom panel) 282

Figure 8-12. Relative contribution of inputs for all four models, sin/cos and VI are the sum of sine and cosine inputs and the sum of all RS vegetation indices inputs, respectively.. 283

Figure 8-13. Spatial distribution of the relative contribution of the sum of EVI and NDWI for TDA of 50 ha 283

CHAPTER 1. BACKGROUND AND GENERAL INTRODUCTION

1.1. Background

Watershed disturbances can be classified into two main classes; natural and anthropogenic. Natural disturbances include wildfire, insects, floods, and droughts with wildfires being the most important and the greatest economic burden (Smith et al. 2003). Anthropogenic disturbances are many, however timber harvesting and oil and gas exploration/production are the two sources of greatest economic impact in the Boreal plain of Canada (Smith et al. 2003). Such disturbances can potentially alter the hydrologic budgets and may expose soils to erosion, resulting in the potential for increased export of nutrients and sediment to surface waters (Munn and Prepas 1986; Cooke and Prepas 1998; Smith et al. 2003). The resulting excessive nutrient loads can cause an imbalance in biomass production in an aquatic ecosystem. The system can then react by producing more phytoplankton than can be consumed by the ecosystem. The resulting biomass overproduction can lead to a variety of problems ranging from anoxic waters (through decomposition) to toxic algal blooms and a decrease in habitat diversity, and thus leading to habitat destruction (Hallengraeff 1993; Chorus 2001; Landsberg 2002). Algal blooms' impacts adversely affect not only the health of people, animals, and aquatic organisms, but also the "health" of local and regional economies (Hoagland et al. 2002). Hence, nutrient modelling and, in particular, phosphorus (P) modelling—as a limiting nutrient—is critical to provide the necessary information for responsive watershed management practices.

Numerous conceptual and data-driven models have been developed to simulate flow (Q) and water-phase P dynamics. Section 1.2 summarizes some of the currently available Q and P models, highlighting their advantages and limitations.

1.2. Watershed Modelling

Models of watershed hydrology and water quality parameters are important tools for hydrological investigations for both operational and research purposes. A multitude of applications (e.g., streamflow and water quality parameters forecasting in time and space, the evaluation of the impact of different forest management and agricultural activities on water quantity and quality, and the evaluation of watershed responses to different climate change scenarios) have contributed to the development of a vast number of watershed models starting in the early 1960s (Wagener 2005).

These models are usually a mixture of linear and non-linear functions, combined to represent those processes occurring in a specific watershed and important for the study objectives at hand. Different model classification systems can be recognized as discussed in Section 1.2.1.

1.2.1. Watershed models classification

Watershed models classification can be represented either according to the degree of physical conceptualization or the spatial resolution of the watershed inputs.

Watershed models can be classified based on the spatial resolution into three classes: (1) lumped models that use average values of input variables over the entire watershed area, and thus having minimal data requirements; (2) semi-distributed models that divide the watershed into sub-watersheds, in which each sub-watershed carry distinct set of input variables; and (3) distributed models that are pixel-based in terms of inputs representations and parameter routing, and therefore having huge data requirements. Although using distributed models is conceptually appealing, the superiority of the more complex distributed models over the simpler lumped models is still an issue of debate (Wilcox et al. 1990; Michaud and Sorooshian 1994; Hauhs et al. 1996; Donnelly-Makowecki and Moore 1999; Gan et al. 2006).

On the other hand, it is common to classify watershed models according to the physical conceptualization into three distinct types: (1) empirical (also called data-driven or black-box); (2) conceptual (also called parametric or gray box); and (3) physically-based (also called mechanistic or white box) model structures.

Empirical models use available time-series of input and output variables (nutrient, precipitation, streamflow, temperature etc.) to derive both the model structure and the corresponding parameter values. They are therefore purely based on the information retrieved from the data and generally do not include, explicitly, prior knowledge about the physical, chemical, hydro-morphological, and biological processes controlling flow processes and contaminant transport mechanisms, hence the name black box. Popular examples of empirical models are Artificial Neural Networks (ANNs) (e.g., Lek et al. 1996; Dawson and Wilby 2001) and time series modelling (e.g., Abrahart and See

2000). Empirical models are usually spatially lumped. A variant on purely empirical models are data-based mechanistic models, that constrain the degrees of freedom to those structures that are physically interpretable, thereby using the hydrologist's understanding of the system under study (e.g., Young 1992; Jakeman and Hornberger 1993; Young and Beven 1994).

Conceptual models are formulated from storage elements that represent parts of the watershed where water is temporarily stored; e.g., soil, aquifers or streams. These elements are filled through fluxes such as rainfall, infiltration or percolation, and emptied through evapotranspiration, runoff, and drainage. The modeler, based on a conceptualization of the watershed, specifies the structure of these models in advance. These models are usually spatially lumped but can be semi-distributed (Boyle et al. 2001). The modeler has to use observations of the watershed response to find appropriate values for the model parameters through empirical relations. Conceptual models form the large majority of models used in practice. Their dependence on measurements of the watershed response (mainly streamflow) limits their use in ungauged watersheds. However, research is ongoing to resolve this problem (e.g., Wagener et al. 2004).

Physically-based models are based on the conservation of mass, momentum and energy. The spatial discretization applied is usually based on grids of pixels (i.e. distributed models), but sometimes also on some type of hydrologic response unit or triangular irregular networks (Beven 2000). The initial philosophy of this class of models was that the degree of physical realism on which these models are based on would be sufficient to relate their parameters to physical characteristics of the catchment under study, thus eliminating the need for observed system response to condition the parameters of the model. However, the currently available physically-based models do not fulfill this ideology. They suffer from extreme data demand, and over parameterization (Beven 1989). The expectation that these models could be applied to ungauged catchments has therefore not been fulfilled (Refsgaard and Knutsen 1996). Beven (1989) argues that this type of models is applied in a way similar to lumped conceptual models, though at a different scale.

1.2.2. Conceptual Watershed Q and P Models

Conceptual watershed-scale Q and P models include; but are not limited to; soil and water assessment tool (SWAT) developed by Arnold et al. (1998), aerial non-point source watershed environment simulation-2000 (ANSWERS-2000) (Bouraoui and Dillaha 1996; Beasley et al. 1980), hydrologic simulation program fortran (HSPF) (Johanson et al. 1984), erosion productivity impact calculator (Sharpley and Williams 1990), annualized agriculture non-point source pollutant loading model (AnnAGNPS) (Bingner et al. 2001), and the Guelph model for evaluating the effects of agricultural management systems on erosion and sedimentation (GAMES) by Cook et al. (1985). The use of this class of models presents the challenge of estimating or calibrating a large number of model parameters from limited available information. Obtaining necessary information for model calibration is always time consuming and expensive.

1.2.3. Data-driven Q and P Models

Data-driven models have been successful in capturing patterns in data with less knowledge of the behavior of the system in terms of interactions between biological, geological, chemical, and physical processes affecting the modelled system and are consequently attractive alternatives to traditional conceptual models. Among those techniques, artificial neural network (ANN) and time series (TS) models hold promise for water quantity and quality modelling.

The application of ANNs in hydrological modelling has been the topic of over 300 refereed publications in the last two decades. The ASCE task committee (2000a,b) and Maier and Dandy (2000) published comprehensive reviews of pertinent work prior to the late 1990s. Most researchers used either a feed-forward multi layer perceptron (FF-MLP) ANN or a recurrent neural network (RNN) for modelling daily streamflow. However, the feed-forward MLP trained with the error backpropagation (BP) algorithm was by far the most widely used network architecture and training algorithm (Castellano-Mendez et al. 2004; Anctil and Tape 2004; Anctill and Rat 2005; Riad et al. 2004; Agarwal and Singh 2004; Riad and Mania 2004; Baratti et al. 2003; Kisi 2004; Tokar and Markus 2000). In all reviewed cases, a sliding window of rainfall and recently observed flow values were utilized to forecast future flow values.

Although ANN has found wide applications in water quality modelling in the last decade (e.g., Brion and Lingireddy 2003; Maier et al. 1998; Maier and Dandy 1996; Wilson and Recknagel 2001; Moatar et al. 1999; Zhang and Stanley 1997; Zhang et al.

2004), ANN modelling of phosphorus dynamics has been limited to a few efforts (e.g., Lek et al. 1996; Holmberg et al. 2006).

In contrast, TS modelling applications in water resources and environmental engineering focused mainly on streamflow forecasting (e.g., Padilla et al. 1996; Abrahart and See 2000; Ali and Dechemi 2004; Chibanga et al. 2003; Hsu et al. 1995; Malgras and Debouzie 1997; Novotny and Zheng 1989; Weeks and Boughton 1987).

Owing to the complexity of hydrologic processes—especially in a forested ecosystem—(Burton et al. 2003; Hewlett 1982), and motivated by the ability of ANNs to model complicated non-linear relationships, almost all previous efforts that compared ANN models to either TS models or conceptual models demonstrated that ANN models performed at least comparatively to, if not better than, other conceptual and statistical models. The focus of this thesis is limited to TS and ANN modelling.

Most of the currently available models, both conceptual and data-driven, are undermined in practice because of the extensive requirement for landscape data (e.g., soils, vegetation, and precipitation) needed for model calibration (Liu et al. 2006; Hauhs et al. 1996; Haan 1989). However, satellite remote sensing (RS) has recently made available a large inventory of cost-effective landscape data over the entire landbase, rather than providing only a sampling of it as is the case with ground-based measurements. In particular, the Moderate-resolution Imaging Spectroradiometer (MODIS) launched by the National Aeronautics and Space Administration (NASA) in December, 1999 has provided scientists with the ability to measure forest growth and snow cover with greatly improved spatial and temporal resolution. Furthermore, the

MODIS data is freely available, thus providing a means of acquiring time series representation of vegetation dynamics and snow cover at an affordable cost. Developing a class of models that can rely on such RS information can potentially boost the applications of watershed water quantity and quality models. Section 1.3 reviews the pertinent literature on the usage of RS information in watershed modelling.

1.3. Application of Remote Sensing in Watershed Modelling

The current resurgence in earth-observing satellite and airborne platforms, along with the advancements in computer and software technology, has made it possible to evaluate and quantify large numbers of watershed physical characteristics and state variables via RS. RS techniques have expanded widely, to the point that they now include most of the electromagnetic spectrum. Different sensors can provide unique information about properties of the surface or shallow layers of the Earth. For example, measurements of the reflected solar radiation give information on albedo, thermal sensors measure surface temperature, and microwave sensors measure the dielectric properties and hence, the moisture content, of surface soil or of snow. This continued development has added new techniques that hydrologists and watershed modelers can use in a large number of applications (Schultz and Engman 2000).

The application of RS imagery can be divided into three main categories: (1) to delineate surface features, such as snow-covered areas, surface water extent or sediment plumes; (2) to retrieve information such as land cover, geological features, or other hydrologic parameters through interpretation and computer classification of remotely sensed data; and (3) to directly use remotely sensed digital data to estimate hydrological state parameters. The third application is the most important to watershed modelling and is normally achieved through electro-optical or statistical modelling of known hydrometric data with satellite data. A thorough description of different applications of RS information in hydrology is summarized in the literature by a number of researchers (Kite and Pietroniro 1996; Pietroniro and Leconte 2000; Pietroniro and Leconte 2005; Smith 1997; Shultz and Engman 2000). Despite the ongoing improvement of remote sensing techniques, only very few success stories on the applications of remote sensing in watershed modelling currently exist (e.g., Andersen et al. 2002; Biftu and Gan 2001, 2004; Boegh et al. 2004; Kite 1998; Sandholt et al. 2003).

Although there has been some success in the application of RS data in hydrology, the application of RS information in watershed water quality modelling has been minimal. A successful nutrient model should rely on information regarding the soil and the vegetation nutrient contents, thus, retrieving remotely sensed information that can be linked to soil/vegetation nutrient interactions can aid in formulating relatively accurate and usable nutrient watershed models. RS vegetation indices (VIs) can represent the vegetation health/stress in terms of the vegetation chlorophyll content and the leaf water content (Cheng et al. 2006). Such information can potentially act as surrogates to soil/vegetation nutrient transport and therefore can potentially represent vegetation dynamics in formulating nutrient models.

The Moderate Resolution Imaging Spectroradiometer (MODIS) launched in 1999, onboard the Terra platform is the primary earth observing system (EOS) sensor for providing data on vegetation dynamics. The MODIS instrument includes seven visible and shortwave bands for land surface studies and provides continuous, 1 to 2-day global coverage at 250 m, 500 m, and 1 km spatial resolutions. Repetitive coverage from this radiometrically and atmospherically corrected MODIS data sets can provide useful information to detect vegetation and land cover moisture contents (Zarco-Tejada et al. 2003), thus the focus in this study is to assess the usefulness of MODIS derived vegetation indices in capturing phosphorus dynamics within a watershed model.

1.4. Research Needs

In Canada and elsewhere, the prediction of daily streamflow (Q) and total phosphorus (TP) concentrations is important for evaluating downstream hydrologic impacts, simulating the impact of extreme floods and droughts, evaluating the impact of different climate change scenarios, and thus for safeguarding against any expected adverse consequences. Providing the resources to gauge all watersheds of interest is practically impossible; thus, most of the currently available models for watershed modelling are limited in practice because of the extensive requirement for landscape data (e.g., soils, vegetation, precipitation) needed for model calibration. Therefore, a class of models that can simulate the response of ungauged watersheds with reasonable accuracy is critical to provide the necessary information for responsive watershed management practices.

With the advent of remote sensing (RS) techniques, the availability of high quality time- and space- variant data at an affordable cost has been made real. Developing a class of watershed models, that can utilize this RS information and that is less reliant on ground-based watershed specific measurements, is expected to overcome the drawbacks of more readily available models that are ground based data collection and time intensive. It will provide a substitute approach using inexpensive RS data with few ground truthing requirements that can move this class of models from a research base to possible industrial applications.

This thesis aims at investigating the possibility of developing a modelling approach capable of simulating streamflow and water quality in ungauged and unmonitored watersheds. It focuses on formulating Q and TP models that are only reliant on currently available meteorological information in Canada, as well as public-domain free-of-cost MODIS RS information. The specific objectives are summarized in Section 1.5.

1.5. Research Objectives

In an attempt to construct a set of Q and TP ANN models that can rely on currently available free-of-cost MODIS RS information as well as scattered meteorological measurements' stations, the following main objectives have to be achieved:

(1) to provide guidelines for modelling time correlated variables using ANN;

(2) to compare ANN modelling to a more traditional manipulation like the multivariate TS approach (ARMAX) highlighting the strength and drawbacks of each technique;

(3) to compare the different interpolation techniques for mapping daily rainfall values attempting to incorporate the most reliable rainfall information available from sparse meteorological and fire tower measuring stations in the modelling activity;

(4) to develop an ANN modelling algorithm capable of modelling Q in ungauged watersheds;

(5) to find a reasonable indicator of hydrologic similarity that can guide model transferability;

(6) to devise a robust ANN TP model that can rely on RS information and currently available meteorological information;

(7) to develop a protocol for linking RS data with ANN models;

(8) to assess the usefulness of MODIS vegetation indices in capturing phosphorus dynamics within a watershed; and

(9) to assess the impact of watershed subdivision on model performance.

1.6. Research Contributions

To achieve the above mentioned objectives, a number of water quantity and quality models were devised and applied to a number of watersheds ranging in basin area from 5 to 130 km^2. The protocols used for data pre-processing, model formulation, and model evaluation have presented the following original contributions to the disciplines of hydroinformatics and environmental informatics:

(1) This thesis presents the first effort to compare autoregressive moving average with exogenous input (ARMAX) modelling to artificial neural network (ANN) modelling for TP predictions. Earlier efforts had focused on comparing the two modelling approaches for Q predictions;

(2) the current study established step-by-step guidelines to ANN modelling of time-correlated variables that can account for data hystereses;

(3) it proposes a feed-forward multi-layer perceptron (FF-MLP) modelling algorithm that relies only on low-cost, readily available meteorological data and careful time series manipulation prior to model building for Q predictions, and thus, is suitable for modelling streamflow in ungauged watersheds;

(4) a new remotely sensed hydrologic similarity measure was proposed in this study, and was found to provide a successful indicator of basin similarity;

(5) the current research is the first to attempt to build a model that can rely on a dynamic suite of remotely sensed vegetation indices for predicting the water-phase TP concentration; and

(6) it is also the first to address the impact of watershed subdivision on a water quality parameter using an ANN modelling algorithm.

1.7. Thesis Organization

In order to preserve the diversity of the models and applications developed in meeting the abovementioned research objectives, a paper format has been employed in preparing this document. Chapter 2 compares ARMAX to ANN modelling of TP concentrations highlighting the strength and drawbacks of each modelling technique. It provides step-by-step guidelines to ANN modelling of time-correlated variables that can account for data hystereses.

Chapter 3 is devoted to modelling the daily change in TP concentration (ΔTP) using ANN. In addition, the impact of wetland area and composition on model formulation and performance was also assessed.

The struggle to develop better models of hydrologic processes requires special attention to data quality. Rainfall time series, in particular, is an important input in hydrologic water quantity and quality models. Kriging is now commonly used as a mapping technique; however, practitioners often get confused in the realm of kriging methods and other interpolation schemes available. In Chapter 4, five geostatistical interpolation techniques; simple kriging (SK), ordinary kriging (OK), multi-Gaussian kriging (MGK), log-normal kriging (LNK), and kriging with an external drift (KED), in addition to the more traditional inverse distance weight (IDW) interpolation technique; were compared for their merits in mapping daily rainfall values, attempting to achieve the most accurate rainfall time series to be used for the subsequent devised models. In addition, sequential Gaussian simulation (SGS) was then implemented to produce 100 equiprobable maps of daily rainfall values. A multi-objective approach; that is based on overlaying the map of the kriging variance, the digital elevation model (DEM), and land use/land cover and road networking maps in a geographic information system (GIS) framework to identify the areas of commonly favorable features; was also proposed to identify potential future sampling locations.

Chapter 5 presents the first step towards a generic ANN modelling algorithm for dynamic predictions of daily streamflow in ungauged watersheds. It also proposes a new measure of hydrologic similarity that can be remotely sensed.

The previous efforts on modelling water-phase TP concentrations showed that there is a need for incorporating information regarding soil/vegetation phosphorus content to reasonably model TP concentration. In absence of this information, only the daily change in TP concentration can be accurately predicted. In Chapter 6, we attempted to model TP concentration through a train of models, in which a streamflow ANN model was developed to predict Q from meteorological information. Modelled Q values were then coupled with weather information to serve as inputs for a total suspended solids (TSS) ANN model. Finally, modelled Q and TSS were used to augment weather data in predicting TP concentration.

Chapter 7 describes an attempt to build an ANN model that can rely on a time series of remotely sensed vegetation indices (VIs) for predicting the dynamics of water-phase TP concentration. We examined the possibility of using five literature based VIs; enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), greenness fraction vegetation index (GFVI), simple ratio vegetation index (SRVI), and normalized difference water index (NDWI), in addition to two indices proposed in this study (SRVIm and GFVIm); to provide sufficient landscape information for water-phase TP modeling.

In Chapter 8, the impact of watershed subdivision on TP model predictions was assessed. Automatic delineation of the watershed into different sub-divisions was achieved using a 30 m x 30 m resolution DEM.

Finally, pertinent conclusions from all modelling efforts, an overall assessment of the research program, and recommendations for future work are presented in Chapter 9.

1.8. References

Abrahart, R.J., and See, L. 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol. Process. **14**(11-12): 2157-2172.

Agarwal, A. and Singh, R.D. 2004. Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. Water Resour. Manag. **18**: 285-300.

Ali, T.B., and Dechemi, N. 2004. Daily rainfall-runoff modelling using conceptual and black box models; testing a neuro-fuzzy model. Hydrol. Sci. J.-J. Sci. Hydrol. **49**(5): 919-930.

Anctil, F., and Tape, D.G. 2004. An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. J. Environ. Eng. Sci. **3**: S121-S128.

Anctill, F.and Rat, A. 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. **10**(1): 85-88.

Andersen, J., Dybkjaer, G., Jensen, K.H., Refsgaard, J.C., and Rasmussen, K. 2002. Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. J. Hydrol. **264**(1-4): 34-50.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R. 1998. Large area hydrologic modeling and assessment: Part 1. Model development. J. Am. Water Resour. Assoc. **34**: 73–89.

ASCE Task Committee. 2000a. Artificial neural network in hydrology – I: preliminary concepts. J. Hydrol. Eng. **5**(2): 115-123.

ASCE Task Committee. 2000b Artificial neural network in hydrology – II: hydrologic applications. J. Hydrol. Eng. **5**(2): 124-137.

Baratti, R., Cannas. B., Fanni, A., Pintus, M., Sechi, G.M. and Toreno, N. 2003. River flow forecast for reservoir management through neural networks. Neurocomputing. **55**: 421-437.

Beasley, D.B., Huggins, L.F., and Monke, E.J. 1980. ANSWERS: a model for watershed planning. Trans. ASAE **23**: 938–944.

Beven, K.J. 1989. Changing ideas in hydrology – The case of physically based models. J. Hydrol. **105**: 157-172.

Beven, K.J. 2000. Rainfall-runoff modeling: The primer. John Wiley and Sons, Chichester, U.K.

Biftu, G.F., and Gan, T.Y. 2001. Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data. J. Hydrol. **244**(3-4): 137-156.

Biftu, G.F., and Gan, T.Y. 2004. A semi-distributed, physics-based hydrologic model using remotely sensed and Digital Terrain Elevation Data for semi-arid catchments. International Journal of Remote Sensing. **25**(20): 4351-4379.

Bingner, R.L., Theurer, F.D., Cronshey, R.G., and Darden, R.W. 2001. AGNPS 2001 Web Site. http://msa.ars.usda.gov/ms/oxford/nsl/AGNPS.html.

Boegh, E., Thorsen, M., Butts, M.B., Hansen, S., Christiansen, J.S., Abrahamsen, P., Hasager, C.B., Jensen, N.O., van der Keur, P., Refsgaard, J.C., Schelde, K., Soegaard, H., and Thomsen, A. 2004. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. J. Hydrol. **287**(1-4): 279-299.

Bouraoui, F., and Dillaha, T.A. 1996. ANSWERS-2000: runoff and sediment transport model. J. Environ. Eng. **122**: 493– 502.

Boyle, D.P., Gupta, H.V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M. 2001. Towards improved streamflow forecasts: The value of semi-distributed modelling. Water Resour. Res. **37**(11): 2739-2759.

Brion, G.M., and Lingireddy, S. 2003. Artificial neural network modelling: a summary of successful applications relative to microbial water quality. Water Sci. Technol. **47**(3): 235-240.

Burton, P.J., Messier, C., Smith D.W., and Adamowicz, W.L. (Editors). 2003. Towards sustainable Management of the Boreal Forest. NRC Research Press, Ottawa, Ontario, Canada. 1039 p.

Castellano-Mendez, M., Gonzalez-Manteiga, W., Febrero-Bande, M., Prada-Sanchez, J. and Lozano-Calderon, R. 2004. Modelling of monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. J. Hydrol. **296**: 38-58.

Cheng, Y.B., Zarco-Tejada, P.J., Riano, D., Rueda, C.A., and Ustin, S.L. 2006. Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. Remote Sens. Environ. **105**(4): 354-366.

Chibanga, R., Berlamont, J., and Vandewalle, J. 2003. Modelling and forecasting of hydrological variables using artificial neural networks: the Kafue River sub-basin. Hydrol. Sci. J.-J. Sci. Hydrol. **48**(3): 363-379.

Chorus, I. (Ed). 2001. Cyanotoxins: occurrence, causes, consequences. Berlin: Springer Verlag.

Cook, D.J., Dickinson, W.T., and Rudra, R.P. 1985. GAMES—the Guelph Model for Evaluating the Effects of Agricultural Management Systems in Erosion and Sedimentation. User's Manual. Tech. Rep. 126-71. School of Engineering, Univ. of Guelph, Guelph, Ont.

Cooke, S.E., and Prepas, E.E. 1998. Stream phosphorus and nitrogen export from agricultural and forested watersheds on the Boreal Plain. Can. J. Fish. Aquat. Sci. **55**: 2292-2299.

Dawson, C.W., and Wilby, R.L. 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geog. **25**(1): 80-108.

Donnelly-Makowecki, L.M., and Moore, R.D. 1999. Hierarchical testing of three rainfall-runoff models in small forested catchments. J. Hydrol. **219**: 136–152.

Haan, C.T. 1989. Parametric in hydrologic modelling. T ASAE. **32**(1): 137-146.

Hallengraeff, G.M. 1993. A review of harmful algal blooms and their apparent global increase. Phycologia **32**:79-99.

Hauhs, M., Neal, C., Hooper, R., and Christophersen, N. 1996. Summary of a workshop on ecosystem modelling: The end of an era?. Sci. Total Environ. **183**: 1–5.

Hewlett, J.D. 1982. Principles of Forest Hydrology. University of Georgia Press, USA. 192 P.

Hoagland, P., Anderson, D.M., Kaoru, Y., and White, A.W. 2002. The economic effects of harmful algal blooms in the United States: estimates, assessment issues, and information needs. Estuaries **25**: 819–37.

Holmberg, M., Forsius, M., Starr, M., and Huttunen, M. 2006. An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams and impacts of climate change. Ecol. Modell. **195**(1-2): 51-60.

Hsu, K.L., Gupta, H.V., and Sorooshian, S. 1995. Artificial Neural-Network Modeling of the Rainfall-Runoff Process. Water Resour. Res. **31**(10): 2517-2530.

Jakeman, A.J., and Hornberger, G.M. 1993. How much complexity is warranted in a rainfall-runoff model? Water Resour. Res. **29**(8): 2637-2649.

Johanson, R.C., Imhoff, J.C., Little, J.L., and Donigian, A.S. 1984. Hydrological Simulation Program-Fortran (HSPF) User's Manual. EPA-600/3-84-066. U.S. EPA, Athens, GA.

Kisi, O. 2004. River flow modelling using artificial neural networks. J. Hydrol. Eng. **9**(1): 60-63.

Kite, G. 1998. Integration of forest ecosystem and climatic models with a hydrologic model. Journal of the American Water Resources Association. **34**(4): 743-753.

Kite, G.W., and Pietroniro, A. 1996. Remote sensing applications in hydrological modelling. Hydrol. Sci. J.-J. Sci. Hydrol. **41**(4): 563-591.

Landsberg, J.H. 2002. The effects of harmful algal blooms on aquatic organisms. Res. Fish. Sci **10**: 113–390.

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus concentration and phosphorus load from watershed characteristics using backpropagation neural networks. Acta Oecol. **17**(1): 43-53.

Liu, Y.B., Corluy, J., Bahremand, A., De Smedt, F., Poorova, J., and Velcicka, L. 2006. Simulation of runoff and phosphorus transport in a Carpathian catchement, Slovakia. River Res. Applic. **22**: 1009-1022.

Maier, H.R., and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. **32**(4): 1013-1022.

Maier, H.R., Dandy, G.C., and Burch, M.D. 1998. Use of artificial neural networks for modelling cyanobacteria Anabaena spp. in River Murray, South Australia. Ecol. Model. **105**: 257-272.

Maier, H.R.and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw.**15**: 101-124.

Malgras, J., and Debouzie, D. 1997. Can ARMA models be used reliably in ecology? Acta Oecol.-Int. J. Ecol. **18**(4): 427-447.

Michaud, J., and Sorooshian, S. 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resour. Res. **30**(3): 593-605.

Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural networks. Application of control and validation data series in Middle Loire River. Ecol. Model. **120**: 141-156.

Munn, N., and Prepas, E.E. 1986. Seasonal dynamics of phosphorus partitioning and export in two streams in Alberta. Can. J. Fish. Aquat. Sci. **43**: 2464-2471.

Novotny, V., and Zheng, S.H. 1989. Rainfall-Runoff Transfer-Function by Arma Modeling. J. Hydraul. Eng.-ASCE. **115**(10): 1386-1400.

Padilla, A., PulidoBosch, A., Calvache, M.L., and Vallejos, A. 1996. The ARMA models applied to the flow of karstic springs. Water Resour. Bull. **32**(5): 917-928.

Pietroniro, A., and Leconte, R. 2000. A review of Canadian remote sensing applications in hydrology, 1995-1999. Hydrol. Process. **14**(9): 1641-1666.

Pietroniro, A., and Leconte, R. 2005. A review of Canadian remote sensing and hydrology, 1999-2003. Hydrol. Process. **19**(1): 285-301.

Refsgaard, J.C., and Knudsen, J. 1996. Operational validation and intercomparison of different types of hydrological models. Water Resour. Res. **32**(7): 2189-2202.

Riad, S., Mania, J., Bouchaou, L. and Najjar, Y. 2004. Predicting catchement flow in a semi-arid region via an artificial neural network technique. Hydrol. Process. **18**(13): 2387-2393.

Riad, S.and Mania, J. 2004. Rainfall-Runoff model using an artificial neural network approach. Math. Comput. Model.**40**: 839-846.

Sandholt, I., Andersen, J., Dybkjaer, G., Nyborg, L., Lo, M., Rasmussen, K., Refsgaard, J.C., Jensen, K.H., and Toure, A. 2003. Integration of earth observation data in distributed hydrological models: the Senegal River basin. Canadian Journal of Remote Sensing. **29**(6): 701-710.

Schultz, G.A., and Engman, E.T. (eds). 2000. Remote sensing in hydrology and water management. Springer, Berlin.

Sharpley, A.N., and Williams, J.R. (Eds.) 1990. EPIC—Erosion/Productivity Impact Calculator: I. Model Documentation, Tech. Bull., vol. 1768. U.S. Department of Agriculture.

Smith, D.W., Prepas, E.E., Putz, G., Burke, J.M., Meyer, W.L., and Whitson, I. 2003. The Forest Watershed and Riparian Disturbance study: a multi-discipline initiative to evaluate and manage watershed disturbance on the Boreal Plain of Canada. J. Environ. Eng. Sci. **2**: S1-S13.

Smith, L.C. 1997. Satellite remote sensing of river inundation area, stage, and discharge: A review. Hydrol. Process. **11**(10): 1427-1439.

Tokar, A.S. and Markus, M. 2000. Precipitation-Runoff modelling using artificial neural networks and conceptual models. J. Hydrol. Eng. **5**(2): 156-161.

Wagener, T. 2005. Watershed Modelling. Water Encyclopedia. John Wiley & Sons, Inc.

Wagener, T., Wheater, H.S., and Gupta, H.V. 2004. Rainfall-runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London, UK.

Weeks, W.D., and Boughton, W.C. 1987. Tests of Arma Model Forms for Rainfall Runoff Modeling. J. Hydrol. **91**(1-2): 29-47.

Wilcox, B.P., Rawls, W.J., Brakensiek, D.L., and Wight, J.R. 1990. Predicting runoff from rangeland catchements: A comparison of two models. Water Res. Resear. **26**(10): 2401-2410.

Wilson, H., and Recknagel, F. 2001. Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. Ecol. Model. **146**: 69-84.

Young, P.C. 1992. Parallel processes in hydrology and water quality: A unified time-series approach. J. of IWEM. **6**: 598-612.

Young, P.C., and Beven, K.J. 1994. Data-based mechanistic modelling and the rainfall-flow nonlinearity. Environmetrics. **5**: 335-363.

Zarco-Tejada, P.J., Rueda, C.A., and Ustin, S.L. 2003. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sens. Environ. **85**(1): 109-124.

Zhang, Q., and Stanley, S.J. 1997. Forecasting raw-water quality parameters for North Saskatchewan River by neural network modelling. Water Res. **31**: 2340-2350.

Zhang, Q.J., Cudrak, A.A., Shariff, R., and Stanley, S.J. 2004. Implementing artificial neural network models for real-time water colour forecasting in a water treatment plant. J. Environ. Eng. Sci. **3**: S15-S23.

CHAPTER 2. ARTIFICIAL NEURAL NETWORKS AND TIME SERIES MODELLING OF TP CONCENTRATION IN BOREAL STREAMS: A COMPARATIVE APPROACH

2.1. Introduction

The Boreal Plain ecozone in the western portion of the Canadian Boreal Forest is experiencing an increase in the frequency and intensity of both natural (e.g., wildfire and insects) and human-induced (e.g., forest harvesting) watershed disturbances, which are thought to cause a measurable increase in water yield, primarily from overland flow (Smith et al. 2003). The typically high phosphorus content of Boreal Plain soils (Evans et al. 2000) and enhanced mobility of soil nutrients after vegetation removal may, in turn, enhance phosphorus export to water bodies. Snow melt and storm events are critical periods for phosphorus migration to receiving water bodies. During these periods, the soil is more susceptible to erosion, thus leading to elevated particulate phosphorus loads to water bodies (Chanasyk et al. 2003; Prepas et al. 2003; Burke et al. 2005). The resultant increase in phosphorus concentrations in water bodies may promote dissolved oxygen depletion, increased cyanobacterial biomass, and even cyanobacterial toxin production, threatening to destabilize the aquatic ecosystem of the region (McEachern et al. 2000; Prepas et al. 2001). Therefore, a reasonable prediction of TP concentration can aid management plans aimed at preserving aquatic ecosystems

---

A version of this chapter has been published. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006. Artificial neural networks and time series modelling of TP concentration in Boreal streams: a comparative approach. *J. Environ.Eng. Sci.* **5**: S39-S52.

in the region.

The hydrologic, biologic, and chemical mechanisms controlling phosphorus transport are well understood at the microscale level. Upscaling to the larger watershed scale through parameterization and calibration techniques is very data-intensive and economically unfeasible in most cases (Maier et al. 1998). On the other hand, data driven models have been successful in capturing patterns in data, with less knowledge of the behavior of the system in terms of the interactions among the biological, geological, chemical, and physical processes affecting the modelled system and are, consequently, attractive alternatives to traditional conceptual models. Among those techniques, artificial neural network (ANN) and time series (TS) models hold promise for water quality modelling.

ANN has found wide applications in water quality modelling in the last decade. Brion and Lingireddy (2003) predicted the peak microbial contamination of *Giardia* and *Cryptosporidium* spp. in Delaware River, USA. Maier et al. (1998) predicted the incidence of the cyanobacterium *Anabaena* spp. in the River Murray, Australia, four weeks in advance with ANN. They also used ANN for 14-day ahead salinity prediction (Maier and Dandy 1996). Other applications included predicting phosphorus concentrations using watershed characteristics of 927 tributary sites throughout USA (Lek et al. 1996), same-day and 30-day-ahead predictions of algal abundance in six freshwater lakes in Japan and Australia (Wilson and Recknagel 2001), pH modelling in the Middle Loire River, France (Moatar et al. 1999), and colour modelling in the North Saskatchewan River (Zhang and Stanley 1997).

In contrast, TS applications in water resources and environmental engineering focused on hydrology, mainly on stream flow forecasting. Padilla et al. (1996) used autoregressive moving average (ARMA) models to model the flow of the Karastic springs in Spain and France. Abrahart and See (2000) compared ARMA to ANN modelling for river flow forecasting and found that ANN outperformed ARMA in their case study. Ahn (1999) used an autoregressive model with a covariate to predict atmospheric phosphorus deposition in South Florida, USA. Ruan and Wiggers (1997) used the autoregressive multivariate TS model to predict the sediment load from flow time series. TS modelling has also been used for other environmental engineering applications, such as the prediction of air pollution levels (Salcedo et al. 1999) and the estimation of the rate of municipal solid waste generation (Katsamaki et al. 1998). El-Din and Smith (2002*b*) used a transfer function TS approach to model the total suspended solids and chemical oxygen demand in the effluent of a full-scale primary sedimentation tank.

The previous ANN modelling efforts, although successful in their case studies, mostly used a rather "hit and miss" approach in model building and parameter optimization. In addition, although the autoregressive moving average with exogenous inputs (ARMAX) has been used extensively to model streamflow, to our knowledge, it has never been used in the literature to model the TP concentration. Thus, the objectives of this study were to provide guidelines for modelling time-correlated variables using ANN, to compare ANN to a more traditional manipulation like the multivariate TS approach (ARMAX) in order to highlight the strength and drawbacks of each technique, and finally to apply these techniques for modelling total phosphorus (TP) concentration in a 2nd order stream watershed on the Boreal Plain.

2.2. Study Area and Database

The study area is the 16 km^2 Willow Creek watershed, northwest of Edmonton, Alberta, Canada (Prepas et al. 2006). The mean, minimum, and maximum daily temperature data were obtained from Environment Canada (Downsview, Ontario). Data from the Whitecourt airport weather station (latitude 54.15° N; longitude -115.78° W) were used because it is the closest station to the study watershed. Mean daily stream flow (Q) and TP concentration data were collected in 2001 and 2002 as part of the Forest Watershed and Riparian Disturbance (FORWARD) Study (see *J. Environ. Eng. Sci.* Volume 2, special issue, 2003 for details). In these years, high flows occurred in late spring and early summer due to snow melt and rain events (Figure 2-1). The hydrograph in 2001, an average year in terms of total annual precipitation, was derived mainly from storm events, whereas that of 2002, a relatively dry year, was a reflection mainly of snow melt. The series of TP concentrations resembles the flow series in its daily fluctuations (Figure 2-1). Both the Q and TP concentration series peak almost at the same time and have proportional values with time reflecting two highly correlated variables.

2.3. Methodology

The TP concentration time series in receiving water bodies of the Boreal Plain is affected mainly by soil phosphorus concentration and the intersection of the water table with surface soil layers. Hence, the TP concentration in water bodies is likely to increase the most during snow melt and storm events. During these periods, the soil is more susceptible to erosion, thus leading to elevated particulate phosphorus loads to the receiving waters (Chanasyk et al. 2003). Knowledge of the soil nutrient concentration, as well as of the phosphorus loads from other sources like air deposition, is crucial to accurately predict TP concentrations in surface waters. Without this information, the focus is necessarily the daily change in TP concentration.

To achieve the study objectives, ARMAX time series and ANN models were constructed to predict the daily change in the TP concentration (ΔTP). The results are displayed in terms of TP concentration by adding the predicted ΔTP to the previous day's TP concentration to improve the visibility of the results. Finally, the two approaches were compared based on statistical measures of goodness-of-fit as well as visual inspection of the predicted versus measured profiles.

In the following sections, TS modelling is introduced, followed by the ARMAX model results. The ANN modelling and the associated results are then discussed. Finally, the merits of the two approaches are discussed, and guidelines for future modelling are provided.

2.4. ARMAX TP Modelling

Multivariate TS analysis requires identifying the dependency of the current value on past values of not only the same series but also on those of the other time series as well. To evaluate the strength of the relationship between the output time series and the potential input time series, cross-correlation analysis was performed to estimate such strength at different time lags. The lags of the input time series that can be significantly correlated to the output time series were then considered as model inputs. Estimating significant correlations requires the evaluation of several statistical functions: the autocorrelation function (ACF), cross-correlation function (CCF), and partial autocorrelation function (PACF). The ACF measures the linear dependence between two points on the same series observed at different times. The CCF denotes the linear dependence of one series on another as time progresses and is usually used to measure the predictability of one series from another. Sample ACF and CCF can be estimated from the modelled data using Equations 1 and 2, respectively (Shumway and Stoffer 2000).

[1] $$ACF = \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

[2] $$CCF = \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(y_t - \bar{y})}{\sqrt{\left[\sum_{t=1}^{n}(x_t - \bar{x})^2\right]\left[\sum_{t=1}^{n}(y_t - \bar{y})^2\right]}},$$

where n is the number of data points, h is the time lag, t is the time, x_t and x_{t+h} denote observations of variable x at times t and $t+h$, respectively, y_t and y_{t+h} represent observations of variable y at times t and $t+h$, respectively, and $\bar{x}$ and $\bar{y}$ are the mean of x and y, respectively. Although the ACF and CCF are important for model formulation, they do not provide all the information needed to choose a candidate TS model order, especially if an autoregressive component is to be modelled (Box and Jenkins 1970). Thus, PACF is required to provide the missing information needed for model development. The Durbin-Levinson algorithm (Durbin 1960) can be used for sample estimation of PACF.

An alternative representation to TS analysis is spectral analysis, which deals with the frequency domain in place of the time domain. Such representation is very important to highlight the dominant frequencies entrenched in the studied series and to distinguish between true variation and variation due to noise. Although this representation was not used in formulating a candidate ARMAX model, it is described here because it was used later to improve the ANN model construction. The variance profile over the frequency (periodogram) is used to estimate the dominant frequency in all studied series. The periodogram and squared coherence are the Fourier transformed version of the sample ACF and CCF, respectively. Equations 3 through 5 evaluate the smoothed periodogram ($\hat{f}(\upsilon)$) estimated from n data points (Shumway and Stoffer 2000). The squared coherence can be estimated from a bivariate extension of Equation 3:

$$[3] \qquad \hat{f}(\upsilon) = \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} [X_c^2(\upsilon_k + l/n) + X_s^2(\upsilon_k + l/n)]$$

$$[4] \qquad X_c(\upsilon_k + l/n) = n^{-1/2} \sum_{t=1}^{n} x_t \cos\bigl(2\pi(\upsilon_k + l/n)t\bigr)$$

$$[5] \qquad X_s(\upsilon_k + l/n) = n^{-1/2} \sum_{t=1}^{n} x_t \sin\bigl(2\pi(\upsilon_k + l/n)t\bigr),$$

where n is the number of data points, t is the time, x_t is the observation of variable x at times t, ν_k is the frequency of interest, L is a smoothing parameter that should be assigned an odd number and should be fairly small relative to n, and X_C and X_S are the cosine and sine Fourier transformations, respectively.

The development of the ARMAX models incorporated in this study can be divided into four stages. First, the patterns of the sample ACF, PACF, and CCF were analyzed and significant model parameters were identified. Second, maximum likelihood was used for model parameter estimation. Third, the first two stages resulted in a set of models that could define the problem of interest, thus, a selection criterion for the produced models was made based on two statistical measures of goodness-of-fit that penalize the error variance by a term proportional to the number of model parameters (Shumway and Stoffer 2000). The multivariate corrected Akaike's information criterion (AIC_c) suggested by Bedrick and Tsai (1994) and the Bayesian information criterion (BIC) recommended by Schwartz (1978) (Equations 6 and 7, respectively) were used as model selection criteria. The AIC_c and BIC statistics are expected to provide adequate model parsimony and to prevent overfitting of the data. Their use acts as dividing the data into a calibration data set (for parameters estimation) and a cross-validation data set (for testing model generality) (Shumway and Stoffer 2000). Finally, model diagnostic checking was conducted by examining model residuals and comparing them to white noise. The ACF of the residuals with the Ljung Box form of the Q-statistic test (Ljung and Box 1978) was used to examine the model residuals. The ASTSA software package (Shumway and Stoffer 2000) was used for parameter estimation, ACF, PACF, CCF, periodograms and squared coherence calculations.

$$[6] \qquad AIC_c = \ln\left[\frac{RSS_k}{n}\right] + \frac{n+k}{n-k-2}$$

$$[7] \qquad BIC = \ln\left[\frac{RSS_k}{n}\right] + \frac{k\log(n)}{n},$$

where RSS_k denotes the residual sum of squares under the model with k parameters, and n is the number of data patterns.

A key assumption in estimating the previously mentioned statistics is that the studied TS should be weakly stationary (i.e., the mean and variance are assumed constant and independent of time, and the covariance is assumed to be a function of the lag h and not the time). This condition is a decision of pooling data together for subsequent statistical calculations. Investigating the ACF and PACF of the Willow Creek TP concentration TS (Figure 2-2) provided sufficient evidence that the stationary condition had been violated. The slow decay in the ACF and the peak at lag 1 in the PACF indicate nonstationary behavior (Shumway and Stoffer 2000). Willow Creek flow TS exhibited a similar behavior. To be able to model these series, a transformation was mandatory to attempt to achieve stationary conditions. The first difference transformation was found to enhance the process towards stationary conditions significantly. Figures 2-3 and 2-4 demonstrate that the slow decay in the ACF of both flow and TP concentration no longer existed and that the lag-1 peak of their corresponding PACF was highly reduced. These are symptoms of weakly stationary TS, making modelling the first difference series by using ARMAX a valid option (Shumway and Stoffer 2000).

Examining the sample ACF and PACF after applying the first difference operator to the series illustrates that ACF cuts off at lag 5 and that PACF cuts off at lag 7 for TP (Figure 2-3), whereas ACF cuts off at lag 6 and PACF cuts off at lag 8 for Q (Figure 2-4). The sample CCF demonstrated a significant correlation between the flow and TP concentration series in the lag range of (0 to 6), with a maximum correlation of 0.55 at zero-lag (Figure 2-5). This result proved that the daily change in the TP concentration was strongly correlated to the daily change in the mean daily Q. The high correlation between the two series suggested that it is possible to model the daily change in the TP concentration (ΔTP) from Q data by using the ARMAX modelling approach. The results of the smoothed periodograms of both the Q and TP concentration illustrated that most of the power was at frequencies less than 0.03, corresponding to a period of one to two months (Figure 2-6). This result indicated that monthly variation within the typical annual cycle represented the highest contributor to the variance and should be dealt with as the dominant frequency in the data. A peak at a period in the neighborhood of four days was also significant, but with much less power than that of the monthly variation. The strong coherence at the Q and TP frequencies of interest (squared coherence > 0.95) confirmed the likelihood of modelling ΔTP from the Q series (Figure 2-7).

Following the ARMAX model-building guidelines summarized above and expressed in more detail in Box and Jenkins (1970) and Shumway and Stoffer (2000), and based on the examined ACFs, PACFs, and CCF, we concluded that ΔTP concentration at time t was likely a function of ΔTP concentration for seven lags and ΔQ for six lags. The previous set of possible inputs resulted in 126 different models. All alternative models were formulated, and model parameters were estimated by using maximum likelihood. Both AIC_c and BIC were used for model selection. Equation 8 represents the model that resulted in the minimum value of the two criteria. Equation 8 can be expressed in terms of the TP concentration instead of the ΔTP, as represented by Equation 9.

$$[8] \qquad \Delta TP_t = \beta_1 \Delta Q_t + \beta_2 \Delta Q_{t-3} + w_t$$

$$[9] \qquad TP_t = TP_{t-1} + \beta_1 (Q_t - Q_{t-1}) + \beta_2 (Q_{t-3} - Q_{t-4}) + w_t,$$

where TP_t and TP_{t-1} are the TP concentration in µg L^{-1} at times t and $t-1$, respectively; Q_t, Q_{t-1}, Q_{t-3}, and Q_{t-4} are mean daily stream flows in m^3 s^{-1} for times t, $t-1$, $t-3$, and $t-4$, respectively; β_1 and β_2 are model parameters; and w_t is the model residual at time t. The values of β_1 and β_2 were found to be equal to 204.6 ($P < 0.001$) and 134 ($P < 0.001$), respectively.

Cross-correlation analysis of the air temperature and TP concentration revealed the low dependency of the TP on the air temperature. However, because the air temperature plays an important role in the dynamics of the TP in the environment and because the weather data were already available via Environment Canada, including such information in ARMAX model development was also investigated. The results indicated that adding the air temperature to the model increased both the AICc and BIC, confirming that the previously obtained model (Equation 9) was the most parsimonious.

The model managed to predict the peak locations correctly, but the predicted peak response was overestimated for three peaks (Figure 2-8). During snow melt, ARMAX managed to predict the peaks; however, it tended to oscillate around the mean value between peaks (Figure 2-9). The overall model performance was satisfactory with a coefficient of multiple determination, R^2 (Equation 10) of 0.78. Table 2-1 summarizes the statistical performance of the model results based on five measures of goodness-of-fit.

$$[10] \quad R^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2},$$

where y is the actual value, $\hat{y}$ is the predicted value, and $\bar{y}$ is the mean of the y values.

The final stage of the modelling procedure was to examine the model residuals. The obtained ACF for the residuals resembled white noise with 90% confidence. The Q-statistic test showed a slight deviation from the acceptable chi-squared value at the same confidence level. Attempting to improve model predictions even more, ARMA models were fitted to the residuals. Nevertheless, the model accuracy was not enhanced significantly, and therefore, the obtained model (Equation 9) was represented as the most parsimonious model obtained by using the TS approach.

2.5. ANN TP Modelling

ANN models attempt to implement structures and functions to simulate the data processing capabilities of human brains. The most widespread ANN design consists of an input layer, hidden layer(s), and an output layer, usually termed the multi-layer perceptron (MLP). The input layer introduces inputs to the network and scales them in a range that the neural network can deal with efficiently. The hidden layer(s) processes inputs by applying non-linear activation function(s) attaining the non-linearity of the network. The output layer represents the response of the network (Hayken 1994). A set of connection weights (model calibration parameters), which link neurons together, provides the connection strength between these neurons. Typically, the data to be modelled are split into two sets: a training set for model calibration and a testing set for model validation. ANN models apply a set of rules that change the connection weights iteratively during the learning process until a stopping criterion is reached. Maier and Dandy (1996) and Zealand et al. (1999) reported that over 80% of previous neural network models used a back-propagation (BP) training algorithm, which is a supervised learning technique. In the BP algorithm, the weights initially are assigned arbitrary small values. As training progresses, the weights are updated systematically, and the network output is compared with the target output; meanwhile, the mean squared error (MSE) between the two is calculated. Weight adjustments are made based on an objective function that reduces MSE attempting to reach a global minimum. The training process stops when a prescribed stopping criterion is reached. Performance criteria to measure goodness-of-fit for both training and testing data sets are used to assess model accuracy.

Albeit easy in concept, vigilant care must be taken in the choice of network architecture (number of hidden layers, number of neurons, type of scaling and activation functions, learning and momentum rates, and stopping criteria). A good design should provide parsimony, generalization ability, avoid overlearning (memorizing the data), and should not be stuck in a local minimum.

The ANN modelling approach conducted in this study can be divided into three main phases: data pre-processing, model building, and model evaluation. During the data pre-processing phase, the modeller should develop an understanding of all the data features, identify possible model inputs and highlight their proportional importance, and survey the possible causes of any unexpected feature entrenched in the data. The model building phase relies on the modeller's experience to construct a parsimonious model capable of not only representing the data used in its building but also generalizing to a wider span of data. This phase includes the choice of the optimum ANN architecture, selection of training algorithm, optimal data division for ANN training and testing, and the decision about when to stop training. The model evaluation includes statistically measuring the goodness-of-fit of the developed model, graphically examining the measured versus predicted profiles, and analyzing the model residuals.

As demonstrated by the TS approach, the daily change in the TP concentration is a function mainly of the change in the mean daily Q at different time steps (being the outcome of all hydrologic processes defining the study watershed). Although air temperature did not enhance the ARMAX prediction skill, it was still included in the ANN model attempting to achieve the best possible model using this approach. TP and Q are known to experience hystereses loops due to the seasonal TP/Q cyclic behavior, and thus, spectral analysis was used to identify the predominant data cyclicity. This information was then fed to the ANN in terms of additional input parameters, as described later in this section.

The CCF of Q and TP concentration (Figure 2-5) demonstrated that the daily change in TP concentration was strongly correlated to Q at the same time through six days ahead. Initial model runs showed that the addition of Q inputs beyond three days ahead did not have any impact on ANN predictions, likely because the CCF values above lag three slightly deviated from the 95% confidence boundaries. Hence, four inputs were used to represent Q, and maximum, minimum, and average air temperatures were used to account for the temperature effect. To account for the seasonal variation in TP/Q behavior, spectral analysis was utilized. The motivation for the method was based on the fact that any parameter subject to natural periodic variation can be approximated with sine and cosine waves at different phase angles (Chatfield 2001). The TP/Q hystereses loops was dealt with by introducing two additional model inputs namely, $\sin(2\pi\nu\tau)$ and $\cos(2\pi\nu\tau)$, where ν represents the dominant frequency of the data, and τ is a time index. The dominant frequencies were obtained from the smoothed periodogram of the TP concentration (Figure 2-6). This frequency range corresponds to a time interval of one to two months. One month ($\nu = 1/12$) was used in this study for not missing any period of interest.

Therefore, 10 neurons were used in the input layer (representing Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, TP_{t-1}, T_{min}, T_{max}, T_{avg}, $sin(2\pi\ \tau/12)$, *and* $cos(2\pi\ \tau/12)$), where Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3} represent mean daily stream flow at the downstream end of the creek at times *t*, *t-1*, *t-2*, and *t-3*, respectively; τ is the month number; and T_{min}, T_{max}, and T_{avg} were the minimum, maximum, and average air temperatures, respectively. A linear scaling function was used to scale the inputs in the range of <<-1 , 1>>. Only one neuron was used in the output layer to represent TP concentration. The tanh activation function was used in the output layer because it outperformed linear, sigmoidal, and Gaussian functions. The initial model runs demonstrated that one hidden layer with one activation function was not adequate to represent the complex system under investigation. Increasing the number of hidden layers did not improve prediction, as the network started to memorize the data. Upon examination, three distinct processes (base flow, snow melt, and storm events) became apparent. To account for possible variation in the system response when dominated by any of the three distinct processes, the hidden layer was constructed from three slabs with different activation functions. Different activation functions applied to hidden layer slabs can detect different features in a data pattern processed through the network. Thus, a three-slab hidden layer design may use a Gaussian function on one hidden slab to detect features in the mid-range of the data, and use a Gaussian complement in another hidden slab to detect features from the upper and lower extremes of the data. Using the sigmoidal function in the third slab can also be helpful to map irregularities in the data and patterns not captured by the other Gaussian functions. Thus the network can "view" the data analogously to the three distinct processes. The sigmoidal, Gaussian, and Gaussian complement functions were the selected activation functions for the three slabs because they revealed the highest model performance in terms of R^2. Equations 11, 12, and 13 denote the sigmoidal, Gaussian, and Gaussian complement functions, respectively, representing the hidden layer slabs' transfer functions within the selected network architecture (Figure 2-10).

[11] $$f(x) = \frac{1}{1+e^{-x}}$$

[12] $$f(x) = e^{-x^2}$$

[13] $$f(x) = 1 - e^{-x^2}$$

The data set was divided into two portions, three-quarters for model training (calibration) and one-quarter for model testing (validation). The data set was divided based on an algorithm that targeted a similar frequency distribution of each data set, with extreme and rare values going to the training data set. A back-propagation training algorithm with a batch update technique was used for training. The NeuroShell 2 TurboProp training algorithm was used to achieve this task (available from Ward Systems Group, Inc., USA). In this algorithm, training proceeds through an entire epoch (the number of patterns in the training set) before the weights are updated. It adds all of the weight changes and at the end of an epoch modifies the weights. Training was stopped at the best testing set to prevent overfitting the data. A systematic approach was utilized to choose the optimum number of neurons per hidden layer slab (El-Din and

Smith 2002*a*). Too many neurons can make the network memorize the data, thereby reducing the generalization ability of the model, whereas an insufficient number of neurons can make the network be unable to capture all input/output relations successfully. In this study, a constructive network was used where only one neuron was used per slab, and the network performance was monitored for both the training and testing data sets. We then continued to add neurons and to monitor the network performance (by using R^2) and plotted the R^2 versus the number of neurons for the testing and training data sets. Typically, increasing the number of neurons should enhance the training-set performance, but the testing-set R^2 should increase as long as adding neurons helps in correctly predicting outputs from inputs, and should decrease when the network starts to memorize the data. The use of two neurons per slab achieved the best network performance with the maximum probability of getting a parsimonious model (Figure 2-11). As a final step in model development, scatter plots of the model residuals were examined for possible trends; however, none existed, indicating acceptable model residuals.

The ANN model (Figure 2-10) managed to predict the training and testing data sets with R^2 values of 0.92 and 0.86, respectively. The high R^2 values associated with both data sets provided evidence of good model prediction accuracy and high generalization ability (Figure 2-12). It could be inferred that prediction accuracy was good for the entire data range (all predictions were close to the 45° line). In general, the model was successful in predicting the TP concentration trend, including the low concentrations (mostly corresponding to the base flow) as well as the peaks (corresponding to the rain events and snow melt) (Figure 2-13). Moreover, the model did not suffer from any lag phenomena.

It is important to assess the relative importance of model inputs as described by Garson (1991). Doing so can help improve the data collection protocol, and to allow more time and effort to be expended on more important inputs. Stream flow, as expected, proved to be the most significant input (Figure 2-14). This result supports the hypothesis mentioned previously, that the main source of the TP loading to the stream is the erosion of watershed soils (particulate phosphorus). The periodic nature of the process (seasonal variation impact and TP/Q hystereses) was an important factor, but the temperature effect was the least significant (Figure 2-14).

2.6. Discussion and Implications

The main goal of this study was to compare the merits of two data-driven modelling alternatives for their ability to model the water-phase TP concentration. Thus, each approach was optimized to its best even if the model inputs were different in both cases. However, for the sake of comparison, the times when the TP concentration was estimated to allow ARMAX model formulation were eliminated to provide a similar time basis for comparing both approaches. Table 2-1 summarizes their statistical performance. The devised ANN model outperformed the ARMAX model based on R^2, root mean squared error (RMSE), AICc, and BIC. However, ARMAX produced less bias, reflected by the lower mean relative error (MRE). The better statistical performance, as well as the better match of the measured versus the predicted TP concentration profiles for the devised ANN model, demonstrated that ANN is superior to ARMAX in modelling nonstationary TS and gapped data. The nonlinear processes inherent in the TP dynamics are likely what favored the ANN approach.

Based on our experience and confirmed by the results of this case study, the following guidelines can improve ANN modelling of time correlated variables experiencing hystereses loops:

(1) Model inputs should be divided into causal inputs, time-lagged inputs, and inputs that represent seasonality. Causal inputs are to be identified based on the physical understanding of the modelled process and according to data availability and economy. TS analysis should be included to identify possible time-lagged inputs. A smoothed periodogram is to be plotted and investigated for the dominant periodic nature. Two additional model inputs are to be included to represent such cyclic behavior. This application would provide the modeller with all relevant model inputs.

(2) All input/output data pairs are to be grouped and then divided in two data sets: training for model calibration and testing for model validation. The division should be made by attempting to match the two data sets' histograms with the extreme values being assigned to the training data set. This process can be performed by sorting all input/output data pairs in ascending order with respect to the modelled parameter, and then extracting one input/output data pair for the testing data set every $n+1$ patterns and leaving the rest for training. Here, n represents the ratio of the training to the

testing data sets (a value of two is commonly used in practice; however, this value should be increased for shorter data spans). The histograms of the two data sets should then be investigated, and any absent data patterns in the data sets should be identified. Redistribution should be iteratively conducted until the two histograms have an optimal match.

(3) The ANN is then to be trained by using the error back-propagation algorithm and according to the description in the "ANN TP Modelling" section.

(4) The modeller should start with the simplest converging model possible (this takes few iterations until convergence is attained). Adding network complexity in terms of the number of hidden neurons, changing the activation function per hidden neuron, and choosing the number of hidden layers should be based on the prediction performance criterion set by the modeller (e.g., R^2).

(5) Finally, several statistics like R^2, RMSE, and MAE, as well as visual inspection of the measured versus predicted profiles and model residuals, should be utilized for model evaluation.

Figure 2-15 summarizes the proposed guidelines for modelling time-correlated variables using back-propagation ANN models.

The current study established step-by-step guidelines for modelling time-correlated variables using ANN. It provided two modelling alternatives that can be applied for the real-time prediction of the TP concentration in the Willow Creek watershed. Such models can also act as transfer functions to simulate the impact of scenario-based daily increases in streamflow on daily changes in TP concentration. These scenarios can be designed to reflect important factors like climate change impacts. Moreover, the algorithm can potentially be generalized for larger industrial landscapes and different snow-pack-dominated constituents.

2.7. Conclusions and Final Remarks

In this study, two approaches were used to model the in-stream TP concentration for the Willow Creek watershed, Alberta, Canada. Based on the modelling results, the following conclusions can be drawn:

- Both ARMAX and ANN predicted the TP concentration with reasonable accuracy. R^2 values of 0.78, 0.86, and 0.92 were obtained for ARMAX, the ANN testing data set, and the ANN training data set, respectively.
- Based on 4 measures of goodness-of-fit statistics (R^2, RMSE, AIC_c, and BIC) and by examining the predicted versus the measured TP concentration profiles for the two modelling approaches, it was evident that ANN outperformed ARMAX. The inherent nonlinearity of the modelled process is likely to favour the ANN approach.
- Whereas ANN was capable of modelling gaped data efficiently, the ARMAX approach required equally spaced data values; therefore, values had to be estimated for the times when measurements were not available for model development.

- In this study, a multi-slab ANN was designed and utilized, in which a three-slab hidden layer with three different activation functions was used to reflect the distinct system behavior with respect to base flow, snow melt, and rain events.

- The seasonal variation in the TP/Q behavior was adequately addressed by coupling spectral analysis and ANN techniques.

- The strength of the ANN connection weights can reflect the relative importance of the input parameters. Applying this concept demonstrated that the flow and seasonal indices (representing data periodicity) were more important than temperature variations for predicting TP concentration.

- The current study provided two modelling alternatives that can be applied for the real-time prediction of the TP concentration in the Willow Creek watershed. Such models can also act as transfer functions to simulate the impact of scenario-based daily increases in the streamflow on daily changes in the TP concentration. These scenarios can be designed to reflect important factors like climate change.

- This study also proposed systematic guidelines for modelling time-correlated variables that suffer data hystereses by combining the TS concepts to identify possible time-lagged inputs and the dominant periodicities in the data, and the ANN modelling capabilities. These guidelines can potentially be applied for modelling other water quality parameters including diffuse pollutants associated with the melting of contaminated snow packs.

2.8. References

Abrahart, R.J., and See, L. 2000. Comparing neural network and autoregressive moving average technique for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol. Process. **14**(11-12): 2157-2172.

Ahn, H. 1999. Statistical modelling of total phosphorus concentrations measured in South Florida rainfall. Ecol. Model. **116**: 33-44.

Bedrick, E.J., and Tsai, C.L. 1994. Model selection for multivariate regression in small samples. Biometrics **50**: 226-231.

Box, G.E.P., and Jenkins, G.M. 1970. *Time Series Analysis, Forecasting, and Control.* Holden-Day, Oakland, CA.

Brion, G.M., and Lingireddy, S. 2003. Artificial neural network modelling: a summary of successful applications relative to microbial water quality. Water Sci. Technol. **47**(3): 235-240.

Burke, J.M., Prepas, E.E., and Pinder, S. 2005. Runoff and phosphorus export patterns in large forested watersheds on the western Canadian Boreal Plain before and for 4 years after wildfire. J. Environ. Eng. Sci. **4**: 319-325.

Chanasyk, D.S., Whitson, I.R., Mapfumo, E., Burke, J.M., and Prepas, E.E. 2003. The impacts of forest harvest and wildfire on soils and hydrology in temperate forests: A baseline to develop hypotheses for the Boreal Plain. J. Environ. Eng. Sci. **2**: S51-S62.

Chatfield, C. 2001. *Time Series Forecasting.* Chapman and Hall CRC Press, NY.

Durbin, J. 1960. Estimation of parameters in time series regression models. J. R. Stat. Soc. B. **22**: 139-153.

El-Din, A.G. and Smith, D.W. 2002*a*. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. **36**(5): 1115-1126.

El-Din, A.G. and Smith D.W. 2002*b*. A combined transfer-function noise model to predict the dynamic behavior of a full-scale primary sedimentation tank. Water Res. **36**(15): 3747-3764.

Evans, J.E., Prepas, E.E., Devito, K.J., and Kotak, B.G. 2000. Phosphorus dynamics in shallow subsurface waters in an uncut and cut subcatchment of a lake on the Boreal Plain. Can. J. Fish. Aquat. Sci. **57**(Suppl. 2): 60-72.

Garson, G.D. 1991. Interpreting neural-network connection weights. AI Expert **4**: 47-51.

Hayken, S. 1994. Neural Networks: A Comprehensive Foundation. MacMillian College Publishing, Inc., NY.

Katsamaki, A., Willems, S., and Diamadopoulos, E. 1998. Time series analysis of municipal solid waste generation rates. J. Environ. Eng. **124**(2): 178-183.

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus and phosphorus load from watershed characteristics using backpropagation neural networks. Acta Ecologica **17**(1): 43-53.

Ljung, G.M., and Box, G.E.P. 1978. On a measure of lack of fit in time series models. Biometrika **65**: 297-303.

Maier, H.R., and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. **32**(4): 1013-1022.

Maier, H.R., Dandy, G.C., and Burch, M.D. 1998. Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in River Murray, South Australia. Ecol. Model. **105**: 257-272.

McEachern, P., Prepas, E.E., Gibson, J.J., and Dinsmore, W.P. 2000. Forest fire induced impacts on phosphorus, nitrogen, and chlorophyll *a* concentrations in boreal subarctic lakes of northern Alberta. Can. J. Fish. Aquat. Sci. **57**: 73-81.

Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural networks. Application of control and validation data series in Middle Loire River. Ecol. Model. **120**: 141-156.

Padilla, A., Pulido-Bosch, A., Calvache, M.L., and Vallejos, A. 1996. The ARMA models applied to the flow of Karstic springs. Water Resour. Bull. **32**(5): 917-928.

Prepas, E.E., Pinel-Alloul, B., Planas, D., Méthot, G., Paquet, S., and Reedyk, S. 2001. Forest harvest impacts on water quality and aquatic biota on the Boreal Plain: introduction to the TROLS lake program. Can. J. Fish. Aquatic Sci. **58**: 421-436.

Prepas, E.E., Burke, J.M., Chanasyk, D.S., Smith, D.W., Putz, G., Gabos, S., Chen, W., Millions, D., and Serediak, M. 2003. Impact of wildfire on discharge and phosphorus export from the Sakwatamau watershed in the Swan Hills, Alberta, during the first two years. J. Environ. Eng. Sci. **2**: S63-S72.

Prepas, E.E., Burke, J.M., Whitson, I.R., Putz, G., and Smith, D.W. 2006. Associations between watershed characteristics, runoff, and stream water quality: Hypothesis development for watershed disturbance experiments and modelling in the Forest Watershed and Riparian Disturbance (FORWARD) project. J. Environ. Eng. Sci. **5**(S1): S27:S37.

Ruan, M., and Wiggers, B.M. 1997. Application of time series analysis to urban storm drainage. Water Sci. Technol. **36**(5): 125-131.

Salcedo, R.L.R., Ferraz, M.C.M., Alves, C.A., and Martins, F.G. 1999. Time series analysis of air pollution data. Atmos. Environ. **33**: 2361-2372.

Schwartz, F. 1978. Estimating the dimension of a model. Ann. Stat. **6**: 461-464.

Shumway, R.H., and Stoffer, D.S. 2000. *Time Series Analysis and its Applications.* Springer, NY.

Smith, D.W., Russell, J.S., Burke, J.M., and Prepas, E.E. 2003. Expanding the forest management framework in the province of Alberta to include landscape-based research. J. Environ. Eng. Sci. **2**: S15-S22.

Wilson, H., and Recknagel, F. 2001. Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. Ecol. Model. **146**: 69-84.

Zealand, C.M., Burn, D.H., and Simonovic, S.P. 1999. Short term stream flow forecasting using artificial neural networks. J. Hydrol. **214**: 32-48.

Zhang, Q., and Stanley, S.J. 1997. Forecasting raw-water quality parameters for North Saskatchewan River by neural network modelling. Water Res. **31**: 2340-2350.

Table 2-1. Summary statistics of the two modelling approaches.

	R^2	RMSE (μg L^{-1})	MRE	AICc	BIC
ARMAX model[†]	0.77	36.7	1.3 %	8.22	7.21
ANN model	0.91	23.6	-6.5 %	7.39	6.40

[†] Periods when TP concentration was estimated for the sake of ARMAX modelling were eliminated to compare both models based on similar data patterns.

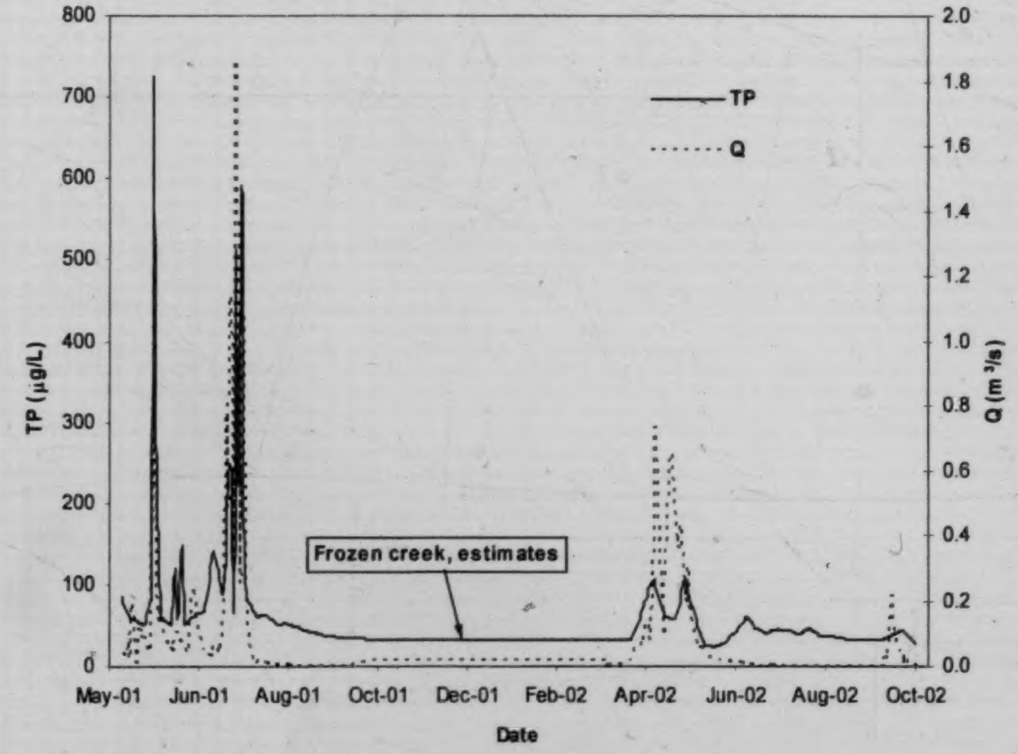


Figure 2-1. Daily stream flow hydrograph and TP concentration profile of Willow Creek watershed from 5 May 2001 through 31 October 2002.

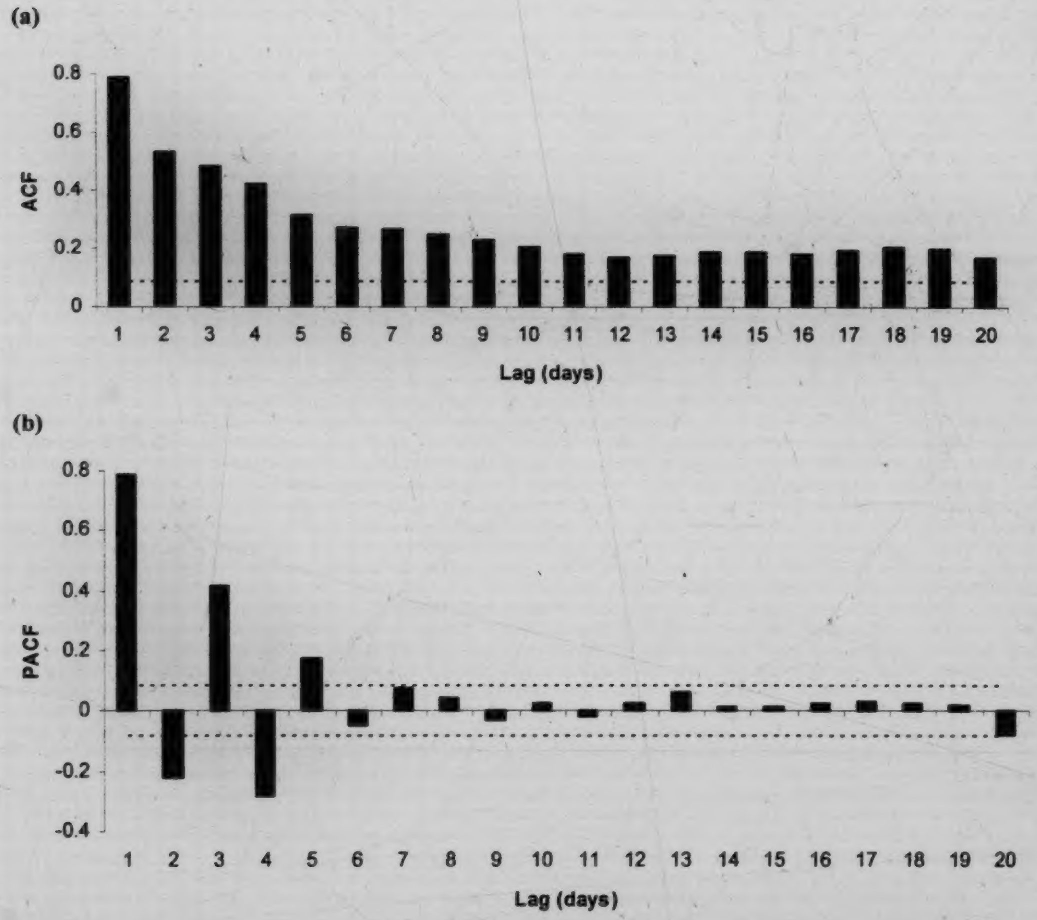


Figure 2-2. ACF (a) and PACF (b) of TP concentration time series, dotted lines showing upper and lower 95% confidence boundaries.

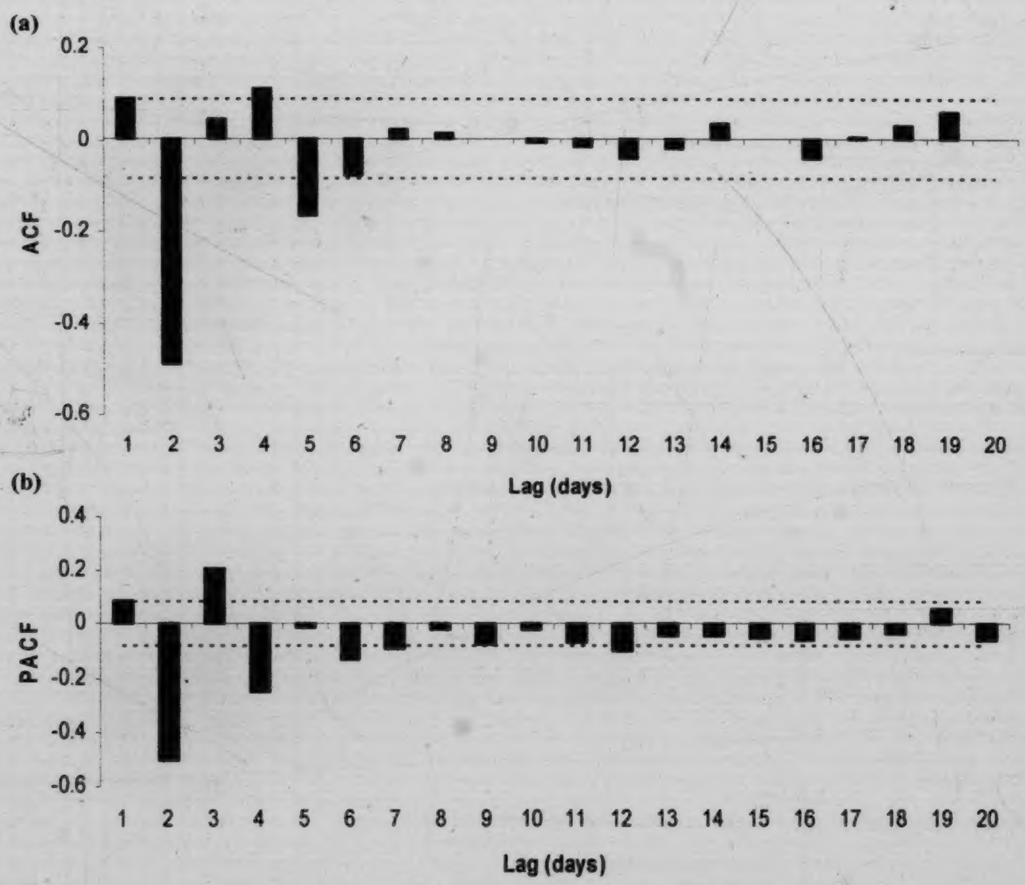


Figure 2-3. ACF (a) and PACF (b) of first difference TP concentration time series, dotted lines showing upper and lower 95% confidence boundaries.

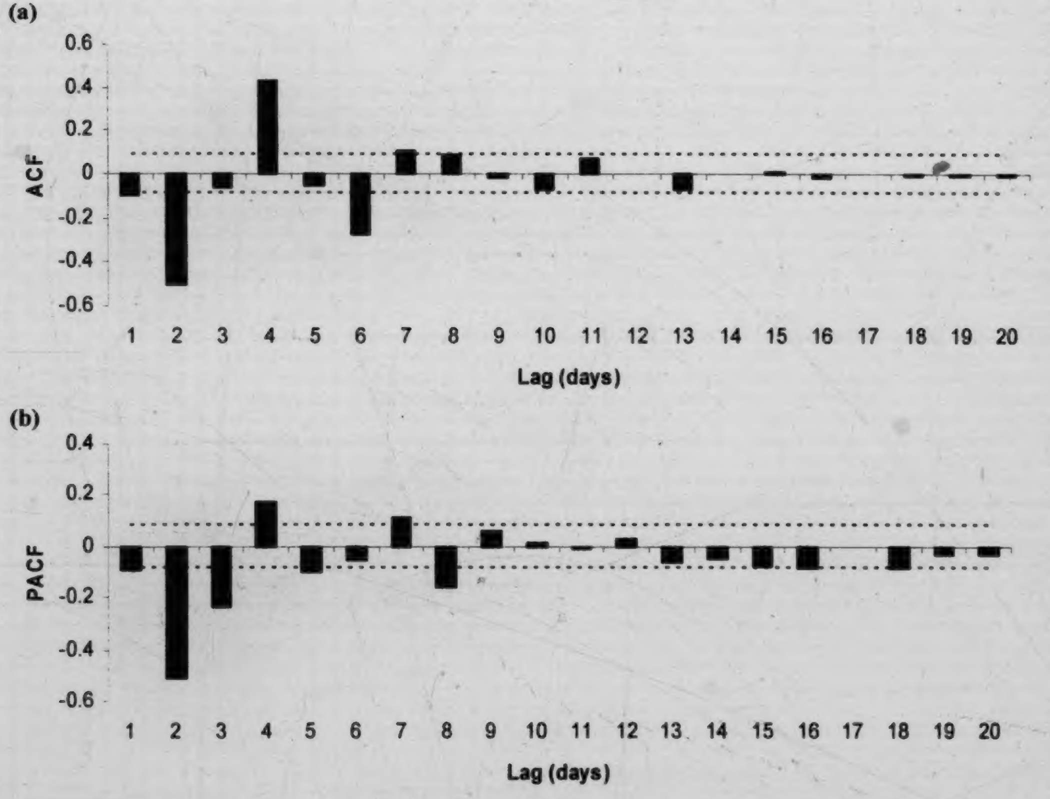


Figure 2-4. ACF (a) and PACF (b) of first difference Q time series, dotted lines showing upper and lower 95% confidence boundaries.

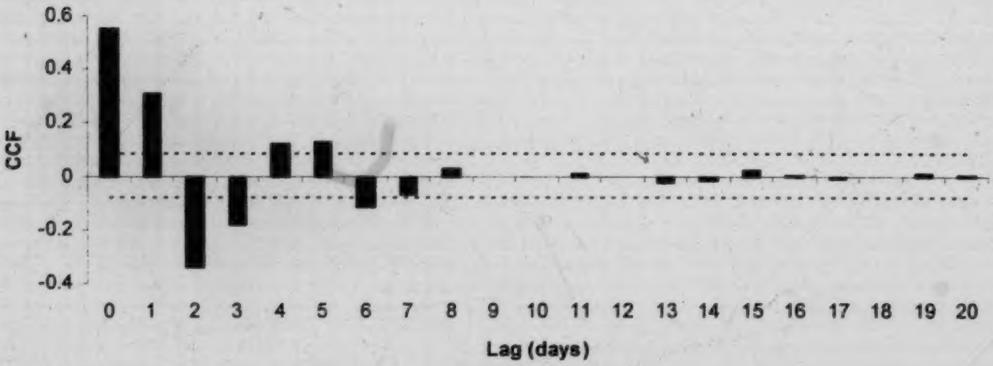


Figure 2-5. CCF of flow and TP concentration time series after applying first difference to both series, dotted lines showing upper and lower 95% confidence boundaries.

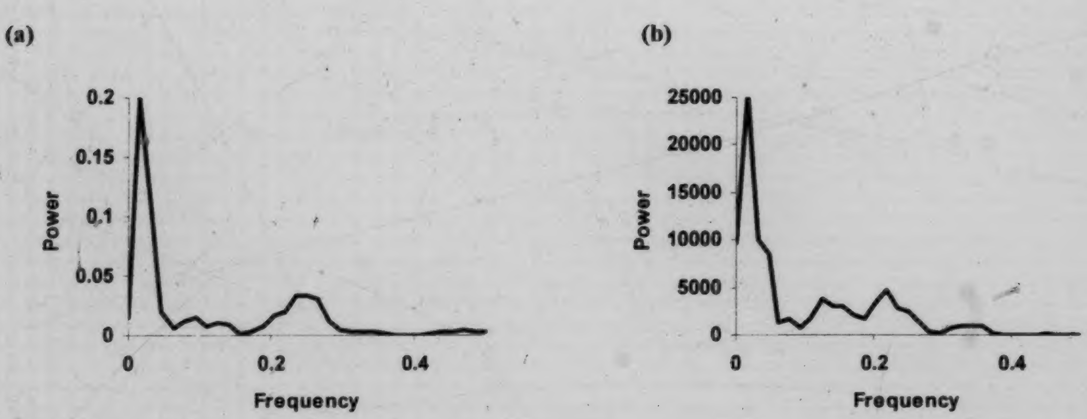


Figure 2-6. Smoothed periodogram ($\hat{f}(\upsilon)$) of Q (a) in ($[m^3/s]^2$ and TP concentration (b) in $[\mu g/L]^2$ time series, L=3.

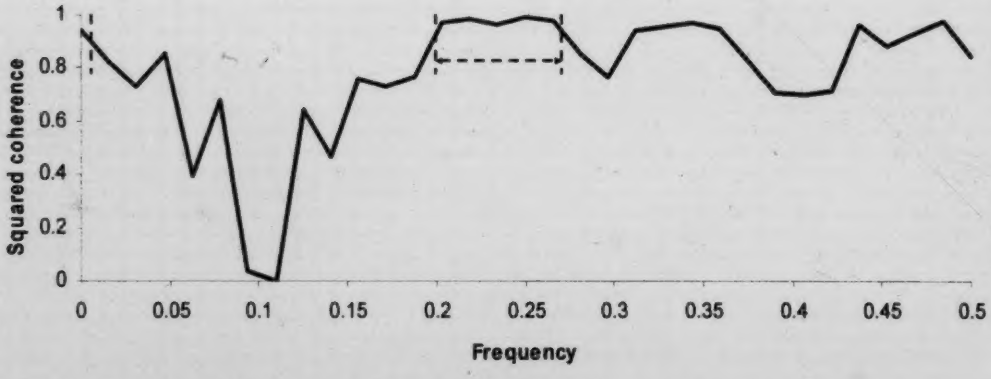


Figure 2-7. Squared coherence of flow and TP concentration time series. Dotted lines depict dominant frequency of both series, L=3.

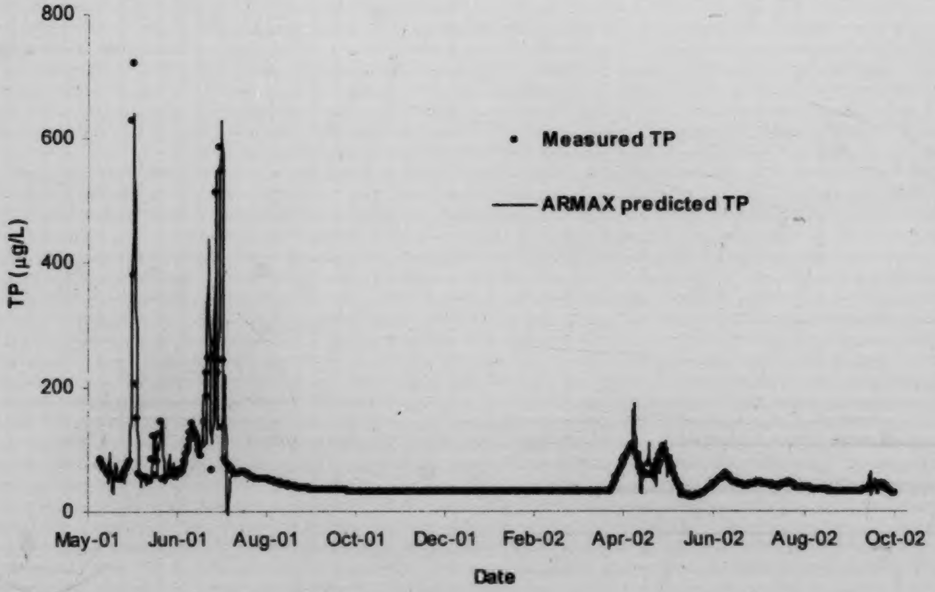


Figure 2-8. Measured and ARMAX predicted TP concentration profiles.

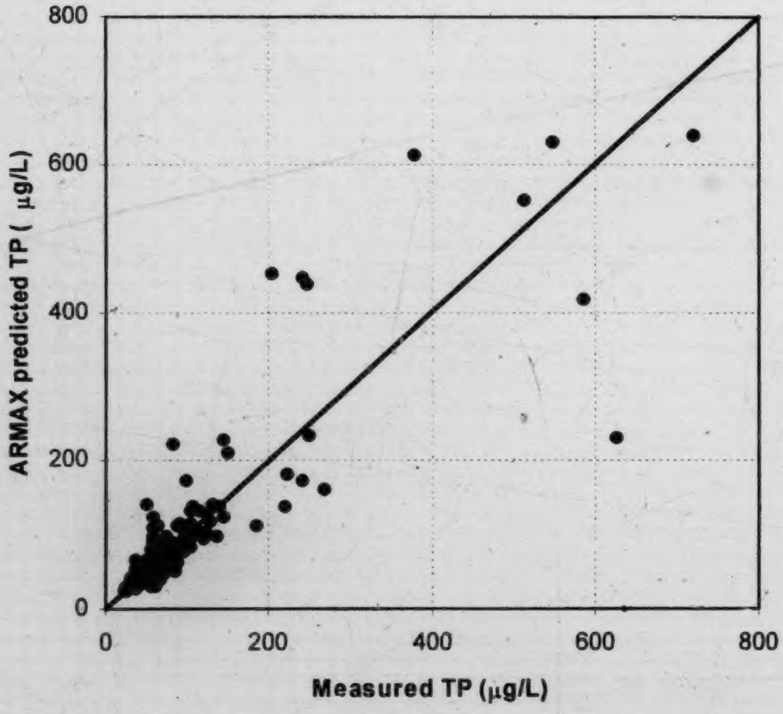


Figure 2-9. Scatter plot of measured and ARMAX predicted TP concentration.

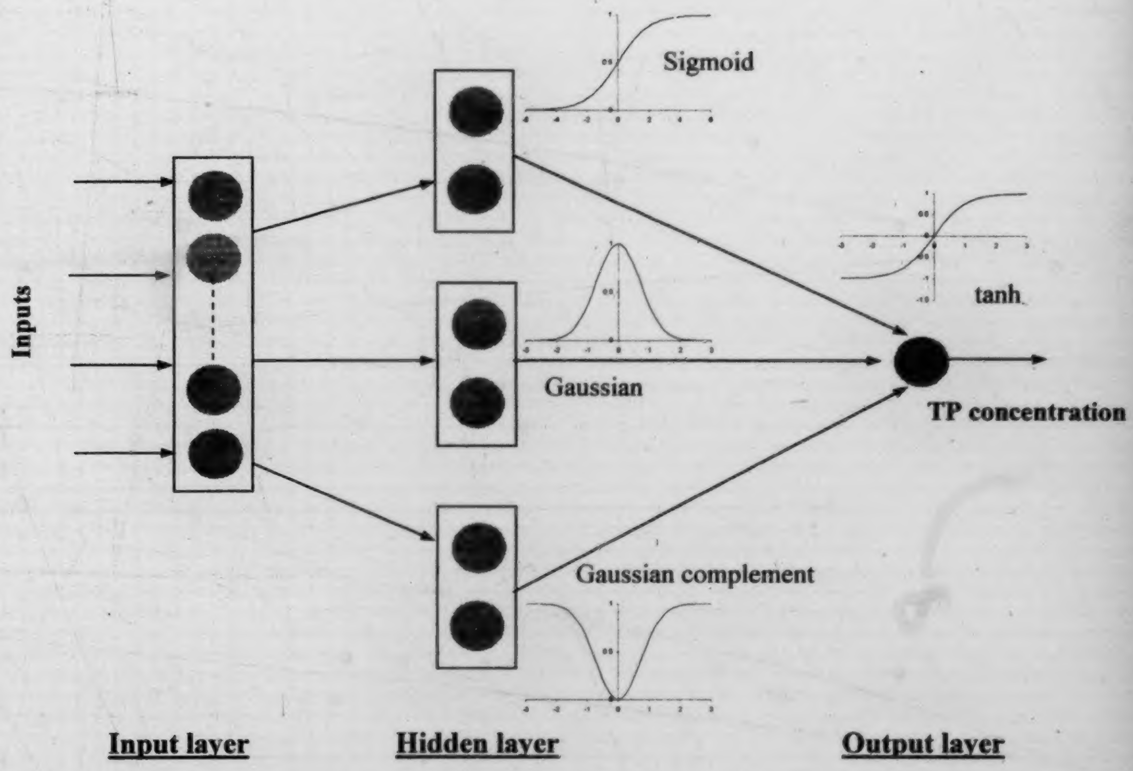

Sigmoid
Inputs
tanh
Gaussian
TP concentration
Gaussian complement
Input layer
Hidden layer
Output layer

Figure 2-10. ANN model selected architecture.

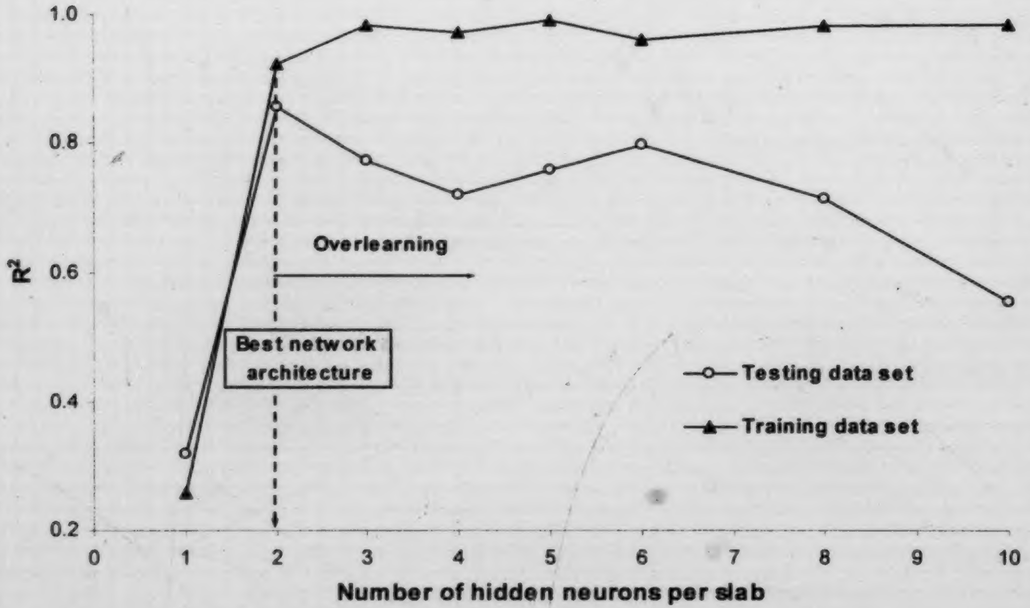


Figure 2-11. Impact of increasing the number of neurons per hidden layer slabs on model performance.

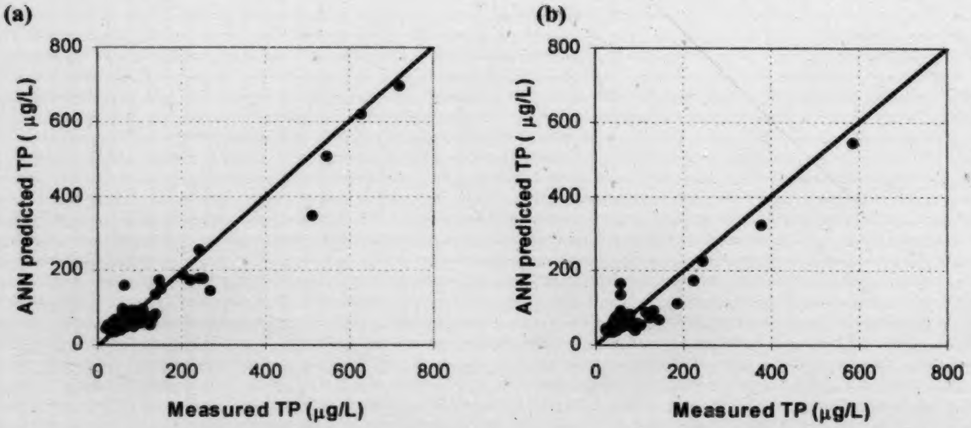


Figure 2-12. Scatter plot of Measured and ANN predicted TP concentration for training data set (a) and testing data set (b).

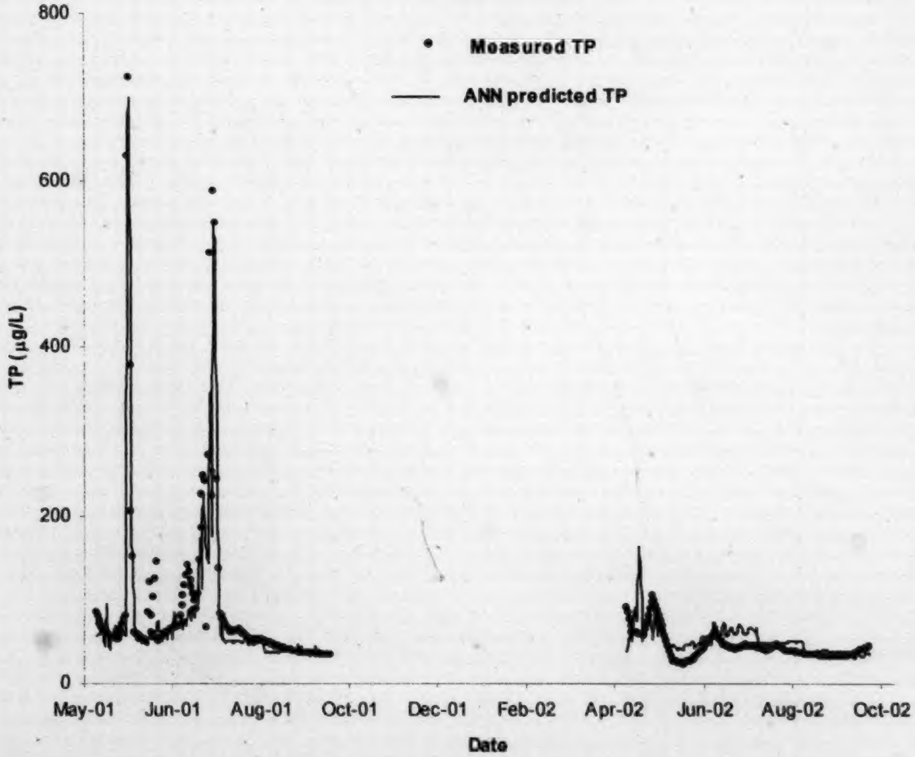


Figure 2-13. Measured and ANN predicted TP concentration profiles.

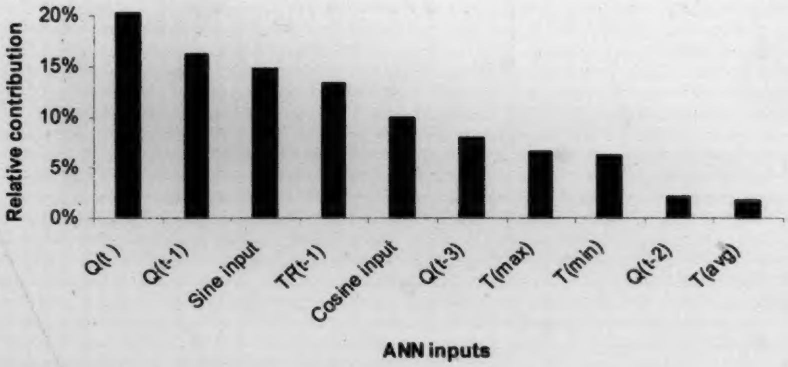


Figure 2-14. Relative importance of ANN model inputs.

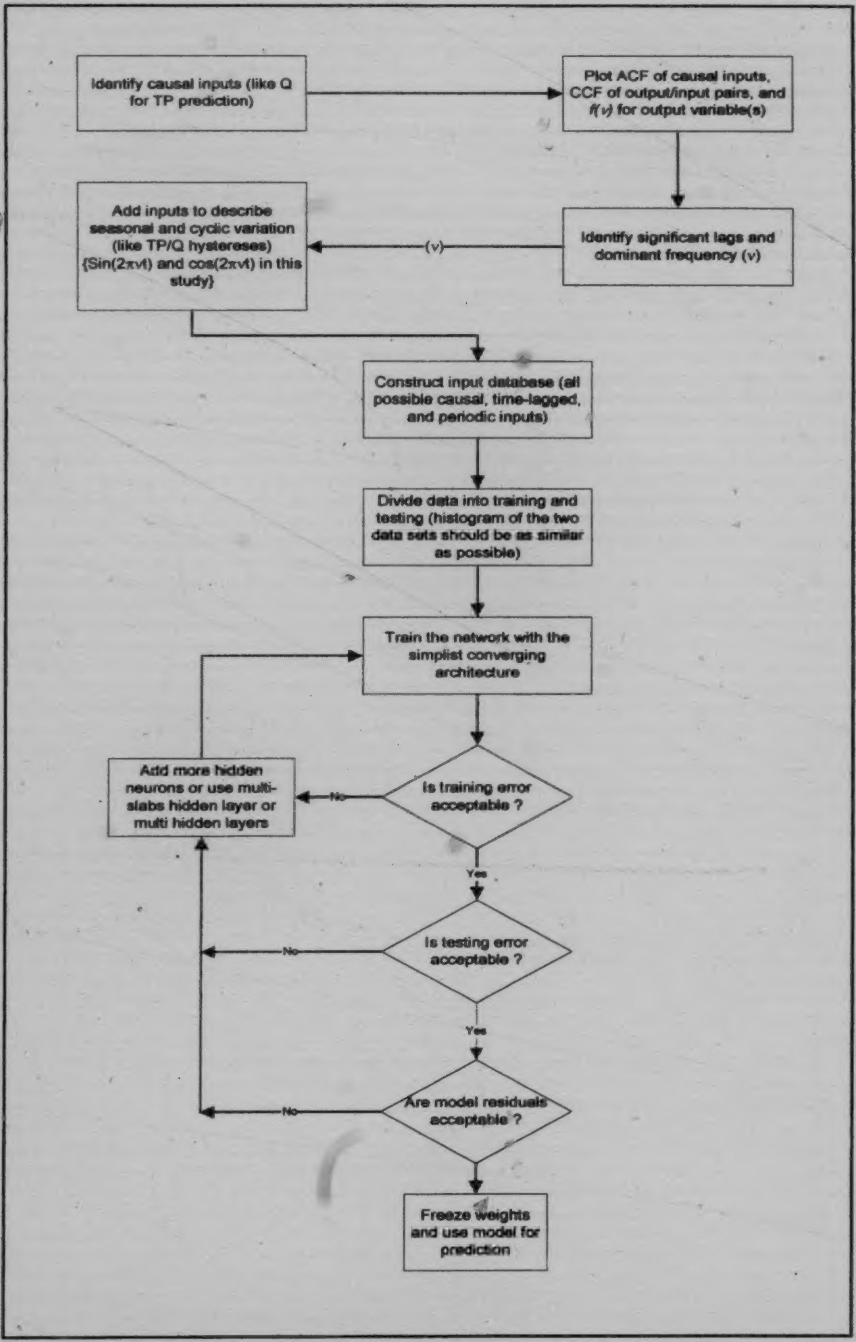

Identify causal inputs (like Q for TP prediction)
Plot ACF of causal inputs, CCF of output/input pairs, and $f(v)$ for output variable(s)
Identify significant lags and dominant frequency (v)
(v)
Add inputs to describe seasonal and cyclic variation (like TP/Q hystereses) {Sin(2πvt) and cos(2πvt) in this study}
Construct input database (all possible causal, time-lagged, and periodic inputs)
Divide data into training and testing (histogram of the two data sets should be as similar as possible)
Train the network with the simplist converging architecture
Is training error acceptable ?
No
Add more hidden neurons or use multi-slabs hidden layer or multi hidden layers
Yes
Is testing error acceptable ?
No
Yes
Are model residuals acceptable ?
No
Freeze weights and use model for prediction

Figure 2-15. Proposed systematic guidelines for modelling time correlated variables using back-propagation ANN.

CHAPTER 3. THE APPLICATION OF ARTIFICIAL NEURAL NETWORKS TO Q AND ΔTP IN SMALL STREAMS ON THE BOREAL PLAIN, WITH EMPHASIS ON THE ROLE OF WETLANDS

3.1. Introduction

Natural (mainly wildfires and insects) and anthropogenic (primarily forest harvesting) watershed disturbance are thought to cause a measurable increase in water yield and water-phase nutrient concentrations in the Canadian Boreal Plain. This in turn has implications for the ecological function in terrestrial and aquatic habitat. Possible consequences of the accelerated rate of watershed disturbance are dissolved oxygen depletion, increased cyanobacteria growth, and cyanobacteria toxin production, threatening to destabilize aquatic ecosystems of the region (McEachern et al. 2000; Prepas et al. 2001). Ecological considerations in forest management in the Canadian Boreal forest have focused on forest polygons and terrestrial biodiversity. The Forest Watershed and Riparian Disturbance (FORWARD) project looks into the problem from a watershed perspective rather than a forest polygon viewpoint. It is a multi-sector interdisciplinary initiative based in the Province of Alberta, Canada. The study area is located in the Virginia Hills, Alberta (Figure 3-1).

A version of this chapter has been published. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006. The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. *Ecol. Modell.* **191**: 19-32.

Extensive monitoring of soil, vegetation, and water phases of 16 small watersheds (3 to 250 km^2) began in the year 2000. Seven of the watersheds are undisturbed systems, four were up to 100% burned during the massive Virginia Hills fire of 1998, whereas four were harvested in the winter of 2003. Stream discharge and water quality data were collected and analyzed for suspended solids, phosphorus, nitrogen, and dissolved organic carbon (detailed description of the project is presented in the special issue of *J. Environ. Eng. Sci.* volume 2, 2003). The area is exemplified by low topographic relief and alkaline phosphorus-rich soils (mainly fine-textured Luvisols) developed from sedimentary bedrock (Prepas et al. 2004). During snowmelt and rain events, when soil is more susceptible to erosion, it represents the highest threat with respect to phosphorus migration to water bodies. Particulate phosphorus, being the dominant phosphorus form in the region during storm flow, is the main contributor to nutrient enrichment of receiving streams. This study focuses on modelling total phosphorus (*TP*) concentration based on data collected from two of the undisturbed watersheds from 2001 to 2003. Earlier studies have demonstrated the importance of wetlands in controlling hydrologic processes and nutrient dynamics in the region (Prepas et al. 2004). The two watersheds (1A Creek, 5.1 km^2 and Willow Creek, 15.6 km^2) reflect variations in wetland percentage areal coverage and wetland composition in the study area (Figure 3-1). 1A Creek differs from Willow Creek not only in the percentage of area occupied by wetlands (30% and 10% for 1A and Willow, respectively) but also in wetland composition. Watershed 1A is distinguished by having over 75 % of its wetland surface area occupied by the so called "toe seepage 2" map unit adjacent to stream channels (Figure 3-2). This map unit shares moist to wet soil conditions of other wetland types. It differs in that forest vegetation is often lacking, and instead these sites are dominated by shrubs, alder and willow. Unlike the other map units, it is not found at major slope inflection points, but occur higher up on the plateau flanks. The physiographic setting and soil conditions within this unit suggest that groundwater discharge from the ground water of neighboring watersheds could be responsible for the wet soil conditions. Just to the west of where these units were mapped lies an area of several square kilometers of muskeg, considerably higher in elevation, which may be the source of the discharge in the 1A watershed. In addition to soil moisture conditions, soil profiles often have organic enriched A-horizons reflecting good conditions for vegetation growth. In addition, the wettest sites typically have peat layers. Common soil profiles include Terric Mesisol in depressional slope positions, Gleysolic (Humic and Luvic Gleysols) at lower to mid slope positions and Luvisolic (Gleyed Dark Gray Luvisolic subgroup) at mid to upper slope positions. The parent materials are primarily moderately fine textured till. These features make this wetland composition distinct and is, thus, suspected of being transiently disconnected from stream networks eliminating and/or delaying the influence of this land area on water and nutrient export, especially in periods when the soil is still frozen early in the snow melting season (Prepas et al. 2004).

The temporal and spatial dynamics of physical, chemical, hydro-morphological, and biological processes controlling phosphorus transport are relatively well understood at the microscale. Upscaling to the larger watershed through parameterization techniques is very data intensive and economically infeasible in most cases (Maier et al. 1998). Scarcity of relevant data at the watershed scale and the heterogeneity and incomplete understanding of biogeochemical processes at this scale make physically based models of limited use (Band et al. 2001). On the other hand, data driven models have been successful in capturing trends, with less knowledge of the behavior of the system in terms of interactions between biological, geological, chemical, hydro-morphological, and physical processes affecting the modeled system. Data driven models are consequently attractive alternatives to traditional conceptual models. Amongst these techniques, artificial neural networks (ANNs) and time series (TS) models hold promise for water quality modelling. The focus of this work is the ANN models due to their capabilities in handling multi-dimensional non-linear problems, their ability to learn from examples, and to generalize to a wider data domain. However, TS analysis is also included for better representation of the modeled processes. In the last decade, ANN has found a wide range of applications in the area of water quality modelling. Examples include: (1) Brion and Lingireddy (2003) predicted peak microbial contamination of *Giardia* and *Cryptospordium* spp. in Delaware River, USA; (2) Maier et al. (1998) modeled the incidence of *Anabaena* spp. in the River Murray, Australia four weeks in advance; (3) Maier and Dandy (1996) used ANN for salinity prediction 14 days in advance; and (4) other applications include: predicting phosphorus concentrations based on watershed characteristics of 927 tributary sites throughout the USA (Lek et al. 1996), same-day and 30-day ahead predictions of algal abundance in six freshwater lakes in Japan and Australia (Wilson and Recknagel 2001), pH modelling in Middle Loire River, France (Moatar et al. 1999), and color in the North Saskatchewan River (Zhang and Stanley 1997).

The objectives of this study were: (1) to provide a modelling tool to predict the average daily flow and the daily change in the *TP* concentration that can be applied for both gauged and ungauged forested watersheds; and (2) to assess the impact of the wetland area and composition on model formulation and performance based on two selected case studies, 1A Creek and Willow Creek watersheds.

3.2. Time Series Analysis

A time series is a collection of samples of a given variable measured at different points in time. Time series analysis typically expresses the information contained in the data in either the time domain or the frequency domain (Shumway and Stoffer 2000).

The time domain analysis is motivated by the assumption that the correlation between adjacent points in time is best explained in terms of a dependence of current value on past values of the same parameter of interest. Multivariate time series analysis not only requires identifying the dependency of current value on past values of the same series but more importantly on other time series as well. To evaluate the strength of the relationship between the output time series and the potential input time series, cross-correlation analysis has to be performed to estimate such strength for several time lags. The time-lagged inputs that can be correlated with the output variable are then considered as model inputs. To identify the significant time-lagged inputs for all four models developed in this study, cross-correlation analysis was adopted. A statistical cross-correlation function (CCF) was estimated for the candidate model output time series and the potential input time series (Shumway and Stoffer (2000) provided a detailed description of this technique). Figure 3-3 presents the CCF of 1A Creek flow and *TP* concentration to demonstrate the approach. It shows that *TP* concentration can be significantly affected by flow for six consecutive lags. The CCF estimation is not very accurate if the time series under study is not weakly stationary, which is the case in the studied series. Hence, the identified significant time-lagged inputs are dealt with as upper boundaries, and the possibility of eliminating some of these inputs is further explored in an attempt to build a parsimonious candidate ANN model.

Frequency domain analysis or spectral analysis assumes that the primary characteristic of interest in any time series relates to the periodic systematic sinusoidal variation found in data. It is inspired by the possibility of representing any complex signal as the sum of pure sine and cosine wave functions. This representation of a time series allows emphasis of the important information within a time series and the opportunity to distinguish between the true variation and the variation due to noise (Masters 1995). A spectral analysis approach is used in this study to identify the dominant periodic nature in the flow and *TP* time series. This information is used to address hystereses between *TP* concentration and flow. The variance profile over the frequency (power spectrum) was used to estimate the dominant frequency in all studied series. Figure 3-4 is a demonstrative power spectrum of 1A Creek flow and *TP* concentration. It shows that most of the power is highlighted at frequencies less than 0.03 (the peak of the power spectrum in Figure 3-4), corresponding to a period of one to two months for both time series. This outcome indicated that the monthly variation within the typical annual cycle represents the highest contributor to the variance and should be dealt with as the dominant frequency in the data.

3.3. Artificial Neural Networks

Inspired by the computational capabilities of human brains, artificial neural networks (ANNs) have found wide applications in recent years. The most widespread ANN design consists of an input layer, hidden layer(s) and an output layer of processing units (neurons). This ANN structure is usually termed the multilayer perceptron (MLP). The input layer introduces inputs to the network, and applies a scaling function to them so that they are in a range that the network can deal with efficiently. The hidden layer then transforms the inputs by non-linear activation function(s), thus attaining the non-linearity of the network. The hidden layer can be divided into multi slabs (each slab is a group of neurons), in which neurons within a slab have the same activation function, whereas neurons in different slabs have different types of activation functions with sparse interconnection between slabs. The output layer represents the response of the network (Haykin 1994). A set of weighted connections link neurons together, reflecting the strength of the connection in magnitude. These connection weights can be looked at as the model calibration parameters. ANNs apply a set of rules that change connection weights iteratively, during the learning process, until a stopping criterion is reached. Maier and Dandy (2000) and Zealand et al. (1999) reported that over 80% of previous neural network models used a back-propagation (BP) training algorithm, which is a supervised learning paradigm. By supervised learning, it is meant that a desired response is available to guide the learning process. In the BP algorithm, the weights are initially assigned arbitrary small values. As training progresses, the mean squared error (MSE) between the target output and the network output is calculated, and the weights are updated systematically. Weight adjustments are made based on an objective function that reduces MSE, attempting to reach a global minimum in the error surface. The training process stops when a prescribed stopping criterion is reached. Figure 3-5 shows the training process, illustrating how the input information propagation and the error back-propagation algorithm are utilized within the neural network architecture.

Typically, data have to be divided into two sets; training and testing data sets akin to the calibration and validation data sets used for conventional model development. However, the authors' experience with ANN modelling suggests that, subject to data availability, data should be divided into three sets in the ratio 3:1:1 for training, testing, and cross validating the model, respectively. The training data set was used to adjust the connection weights. The testing data set allowed building a robust model by determining when to stop training. The ability of the model to generalize (i.e. to produce correct results on previously unseen data, rather than just to memorize the data already encountered during training) was finally measured by applying the developed model to the cross-validation data set. This strategy was adopted by other researchers, for example Maier and Dandy (2000), Ryan et al. (2004), and Zhang et al. (2004).

In this study a MLP ANN trained with the error BP algorithm was used to model both the flow and *TP* concentration of 1A Creek and Willow Creek watersheds. A thorough description of model formulation is given in the model development section.

3.4. Development of ANN Models

This study was aimed at devising an ANN modelling tool that can predict the flow and *TP* concentration for ungauged watersheds (where daily flow is not monitored) in the Boreal Plain of Western Canada. Therefore, all inputs should be easily accessed via a public-domain database, like the Environment Canada weather database, without the need to install flow gauges in each modeled watershed. This requires predicting flow first, then using the predicted daily flow as inputs to predict the associated daily change in *TP* concentration. The ANN modelling approach conducted in this study can be divided into three distinct phases: data pre-processing, model building, and model evaluation.

During the data pre-processing phase, the modeler should develop an understanding of all the data features, identify possible model inputs and highlight their proportional importance, and survey possible causes of any unexpected feature(s) entrenched in the data. The model building phase relies on the modeler's experience to construct a parsimonious model capable of not only representing the data used in its training but also a wider span of data, including that related to other similar watersheds. This includes the choice of the optimum ANN architecture and network internal parameters, the selection of the training algorithm, the optimal data division for ANN training and testing, and the decision about when to stop training. The model evaluation includes statistically measuring the goodness of fit of the developed models, graphically examining the measured versus the predicted profiles, and analyzing the model residuals.

3.4.1. Data Pre-processing Phase

The data pre-processing phase is intended to ensure that all data features are well understood, to identify possible model inputs, and to detect possible causes of any unexpected feature(s) entrenched in the data. Five important features were identified to be consistent in all studied data sets (Willow Creek and 1A Creek flow and *TP* concentration time series):

(1) an annual cyclic nature;

(2) seasonal variations within the year;

(3) the variables are highly correlated with time;

(4) the 2001 flow hydrograph reflects high rain events while both the 2002 and 2003 hydrographs are dictated merely by snowmelt and base flow conditions; and

(5) hystereses loops of flow and *TP* concentration are noted.

Model inputs can be divided into cause/effect inputs, time-lagged inputs, and inputs reflecting annual and seasonal cyclic nature. The cause/effect class of inputs was acquired from an understanding of the physical factors controlling the modeled parameters. This class of inputs is described in more details for the flow models in Section 3.4.1.1., and for the *TP* concentration models in Section 3.4.1.2. However, to address points 1 to 5 mentioned above, time-lagged inputs, as well as inputs reflecting seasonal and annual cyclisity must be identified. To account for the annual and seasonal cyclic nature (points 1 and 2), spectral analysis was conducted to identify the dominant frequency explaining the data. A smoothed periodogram (as the one shown in Figure 3-4) was constructed for all the four series in order to identify the frequencies that contributed most to the variance. The results showed that in an annual cycle, the monthly variation represents the dominant periodicity in the data, and therefore, the dominant frequency (ν) is equal to 1/12. Acknowledging that any periodic function can be approximated by two sine and cosine wave functions at different phase angles (Chatfield 2001), the seasonal variation in the flow and *TP* for the two studied watersheds was represented by adding two distinct model inputs, namely, $\sin(2\pi\nu t)$ and $\cos(2\pi\nu t)$ where t is a time index that varies from 1 to 12 according to the month of the year. Good representation of the periodic features of the data can help the network to differentiate between seasons, to dynamically change the input/output relation according to the season, and thus, to address data hystereses (point 5 mentioned above).

When modelling time-correlated variables, model inputs should not be limited to cause/effect type inputs but should also include time-lagged inputs to address point 3 mentioned above. The cross-correlation time domain analysis, as explained in Section 3.2, was used to identify the possible time-lagged inputs.

3.4.1.1 Flow Models

Traditionally, stream flow has been estimated from rainfall using rainfall-runoff models. As the stream flow is closely related to the amount of rainfall, therefore the measured rainfall time series (R_t) was used as one of the primary model inputs.

Investigation of the modeled hydrographs implied that snowmelt is another important hydrologic process in the studied area. High flow values in early spring were closely linked to snowmelt at this time of each year. The daily snowfall measurements were available from the weather station close to the study area (Figure 3-1 shows the location of the weather station). Other variables controlling water flow include evapotranspiration and soil properties. Measuring these parameters and their variation in space and time is economically infeasible in many cases, and thus, this study attempted to build flow models that utilize limited weather station information, typically available from Environment Canada weather stations scattered on the Canadian Boreal forest. Therefore, the cause/effect inputs used for the flow modelling were rainfall, snowfall, and temperature as indicative of solar energy controlling evapotranspiration and snowmelt in such a forested ecosystem.

Snowmelt is typically estimated by either the energy balance approach or the temperature-index approach. However, due to the difficulty and expense of fulfilling the data requirements of the former, the latter approach is the most extensively used in the literature (Dingman 2002). The temperature-index approach estimates snowmelt as a linear function of average air temperature. The logic relies on the strong correlation between solar radiation and air temperature during snowmelt. This technique equates the daily snowmelt (Δw) as a linear function of the mean air temperature, provided that this temperature exceeds a base temperature for that day (Equation 1). Therefore, during a specified time interval (t), snowmelt is a function of the total of degree-days (dd_t) summed up for days where temperature was above a baseline temperature (typically taken as zero) as represented by Equation 2.

The cumulative snowfall at time t (S_t) was used to reflect the available snow depth, representing the amount of snow available for melting. The cumulative degree-days (dd_t) were used to provide an integrated measure of the heat energy available for snow melting. Thus, cumulative snowfall and degree-days can act as surrogates to the temperature-index snowmelt approach and are therefore used as inputs for the developed flow models.

$$[1] \qquad \Delta w = \begin{cases} f(T_{mean} - T_b), & T_{mean} \geq T_b \\ 0, & T_{mean} < T_b \end{cases}$$

$$[2] \qquad dd_t = \sum_{i=0}^{i=N-1} (T_{mean(i)} - T_{b(i)}).(t_{i+1} - t_i),$$

where, Δw is the daily snowmelt, T_{mean} is the daily average air temperature in °C, T_b is a base temperature typically taken as 0 °C, N is the number of days during which $T_{mean} \geq T_b$, dd_t are the total degree days at time t in °C.day, and $(t_{i+1} - t_i)$ is typically taken as 1 day.

Having identified the cause/effect inputs (rainfall and snowmelt), cross-correlation analysis was used to identify significant time-lagged inputs. Two additional inputs were finally added to address data periodicity and TP/Q hystereses as explained earlier. A summary of the inputs used in this study is presented in Table 3-1.

3.4.1.2 ΔTP models

TP concentration in receiving water bodies of the Boreal Plain is mostly affected by soil phosphorus concentration and the intersection of the water table with surface soil layers. Hence, it is likely that *TP* concentration in water bodies increases maximally during snowmelt and storm events. During these periods, the soil is more susceptible to erosion, yielding elevated particulate phosphorus loads to water bodies. Due to the phosphorus-rich nature of the soil in the studied area (mainly fine-textured Luvisols developed from sedimentary bedrock), it is envisaged that snowmelt and storm events are the driving forces that cause increases in *TP* concentration in water streams (Chanasyk et al. 2003; Prepas et al. 2004). However, knowledge of soil nutrient concentrations, as well as possible phosphorus loads from other sources such as air deposition, is crucial for accurate prediction of water-phase *TP* concentration. In the absence of this information, it is only possible to predict the daily change in *TP* concentration. Nour et al. (2004) developed a model that can predict *TP* concentration using measured flow values. However, continuous flow measurements are not always easy to obtain, so the model's applicability is limited to gauged watersheds. This study provides modelling tools that can be more broadly applied to both gauged and ungauged watersheds. Modeled daily flow values were used as *TP* ANN model inputs instead of flow measurements.

In the case of *TP* modelling, cause/effect inputs were limited to the daily average flow (being the outcome of all hydrologic processes defining the studied watershed), and the daily average air temperature. As with the flow modelling case, the time domain analysis was used to identify possible time-lagged inputs, and the frequency domain analysis was conducted to determine the dominant frequencies embedded in the *TP* concentration time series. However, in this case the daily change in flow (Δ*Q*) rather than the time-variant flow (*Q*) was used as model input because it is the daily change in flow that explains the daily change in *TP* concentration. Table 3-1 summarizes the set of inputs that were used in the final models developed for 1A Creek and Willow Creek watersheds.

3.4.2. Model Building Phase

The model building phase aims at producing a robust ANN model that can accurately map outputs from inputs. A good ANN model should not be confined to mapping the data used during its development, but should be able to generalize by mapping other data sets. Hence, the choice of the training algorithm, the network architecture and internal parameters (number of hidden layers, number of neurons, type of scaling and activation functions, learning and momentum rates, and stopping criterion), and the division of the data into training (calibration), testing (validation), and cross-validation data sets are crucial to achieving a parsimonious model. A single hidden layer with three slabs operating with three different activation functions was found to provide the optimum network architecture for all the devised models. It appears that the three-slab hidden layer can capture the three different hydrologic behaviors of the modeled system (base flow, storm events, and snowmelt). Initial model development attempts demonstrated that one hidden layer with one activation function is not adequate to represent the complex system under investigation. Increasing the number of hidden layers did not improve the models' prediction ability, as the network started to memorize the data. Dividing the hidden layer into three slabs with three different activation functions produced the most accurate model in all the studied cases. The exploration of the studied system highlighted three distinct controlling processes (base flow, snowmelt, and storm events). Interestingly, the three slabs were able to represent the three distinct processes. This modular division in the hidden layer could possibly detect different features of the pattern processed through the network. For example, a network design may use a Gaussian function on one hidden slab to detect features in the mid-range of the data and use a Gaussian complement in another hidden slab to detect features from the upper and lower extremes of the data. The logistic function can also be helpful to map irregularities in the data and extreme patterns that are not captured by the other Gaussian functions. Thus, the network has three ways of viewing the data, analogous to the three distinct processes. The logistic, Gaussian, and Gaussian complement functions were the selected activation functions for the three slabs. The selection was based on the model performance measured by the coefficient of multiple determination (R^2). The optimum neural network architecture for all the four developed models used a linear scaling function that scales the input data in the open interval of -1 to 1, a three-slab hidden layer utilizing the Gaussian, Gaussian complement, and the logistic activation functions, and the output layer consisted of one output neuron with different output activation functions according to the case (Figure 3-5). The choice of the model architecture and internal parameters was judged by the models' performance in terms of R^2 of the testing data set and the ability to recognize all data patterns.

Two training algorithms were tested in this study; the first is the typical gradient descent BP algorithm that utilizes a learning rate and a momentum coefficient to control the training speed and to facilitate moving towards a global minimum in the error surface (Haykin 1994), and the second is a BP algorithm with a batch update technique (BP-BM). NeuroShell 2 software package was used to train the models (Ward Systems 1996). In the batch mode of BP learning, training proceeds through an entire epoch (i.e. it cycles through all of the patterns in the training data set) before the weights are updated. In the BP learning, the repeated training iterations successively increases the performance of the network in the training data set, typically by memorizing the training examples, but the resulting network may perform poorly on other data sets (commonly referred to as network overfitting or network memorization). The methodology adopted to solve this problem was to simultaneously monitor the performance of the training and the testing data sets in terms of R^2. Training continued as long as the error of the testing data set was continuously decreasing and was halted when this error started to increase, even if that of the training data set was still decreasing. However, the connection weights were adjusted only on the basis of the training data set. A systematic approach was utilized to choose the optimum number of neurons per hidden layer slab (El-Din and Smith 2002). The use of too many neurons could cause the network to memorize the data and, thus, reducing the ability of the model to generalize, whilst an insufficient number of neurons could prevent the network from capturing all input/output relations successfully. A constructive neural network was used in which each hidden layer slab was constructed only from one neuron, and the network performance was monitored for both the training and testing data sets. The network performance, in terms of R^2, was monitored as additional neurons were introduced to the hidden layer slabs. The R^2 was then plotted against the number of neurons in the hidden layer for the testing and the training data sets. Typically, increasing the number of neurons should enhance the training-set performance. The testing-set performance would increase whilst the additional neurons help to correctly predict outputs from inputs, and will decrease when the network starts to memorize the data. The advantage of the second algorithm (the BP-BM algorithm) is that it is insensitive to both the learning rate and the momentum coefficient, giving flexibility to less experienced modelers (El-Din and Smith 2002). Table 3-2 summarizes the optimum model architecture and internal parameters utilized for the four developed models in this study.

3.4.3. Model Evaluation Phase

Model evaluation was based on four criteria:

(1) the coefficient of multiple determination (R^2);

(2) the graphical examination of both the measured and the predicted flow hydrographs;

(3) the residuals analyzed and checked for independence, by plotting the residuals versus time and versus flow time series, then the resulting plots were explored for trends; and

(4) the model stability was tested by swapping the testing and cross-validation data sets, then retraining the ANN model and re-assessing the new model performance, a robust model should still perform well, even when the testing and the cross-validation data sets are swapped.

3.5. Results and Discussion

3.5.1. Case study 1: The Willow Creek watershed

The devised flow model for the Willow Creek watershed was able to simulate the average daily flow successfully, with R^2 values of 0.96, 0.85, and 0.84 for the training data set, the testing data set, and the cross-validation data set, respectively. The developed ANN model proved to be stable and consistent in prediction. A high R^2 was retained even by swapping the testing and the cross-validation data sets. The maximum root mean squared error (RMSE) for all the data sets was 0.08 m^3/s. The RMSE was very small compared to the magnitude of the corresponding average daily flow, highlighting the high model performance. The ANN flow model was able to successfully predict both peak flow and base flow and the predictions did not exhibit any lag phenomena when compared to the measurements (Figure 3-6). Residual analysis was then conducted as a final stage in model evaluation. Graphing the residuals is very important in model judgment. If the model fits the data, the residuals should only reflect the measurement error that is assumed to be random. Hence, any lack of randomness in the residuals undermines the strength of the fitted model. In this case, the residuals were plotted versus time and versus flow to check for independence. Residual plots showed only random scatter (no obvious trends were detected, instead, points were scattered all over the graph plot), indicating that the developed models have no serious deficiencies.

The predicted flow values were used to produce inputs for the *TP* concentration model. The statistical performance measures for the devised model are presented in Table 3-3. The R^2 values for the training, testing, and cross-validation data sets were calculated to be equal to 0.95, 0.91, and 0.78, respectively. Although the R^2 value of the cross-validation data set was relatively low, the RMSE ranged from 15 to 34 μg/L for all the data sets. The RMSEs were small in magnitude compared to the respective actual *TP* concentration. The predicted versus the measured *TP* concentration profiles (Figure 3-7) show fair agreement between modeled and measured values for all cases. The residual analysis conducted on model residuals proved to have no structure other than the random structure leaving no statistical reason to question the model's validity.

The performance of both the flow and the *TP* concentration ANN models was high in modelling the Willow Creek watershed. The low percentage of the wetland areal cover (10% of the watershed area) as compared to 30% in case of the 1A Creek watershed did not affect the accuracy of model predictions, and thus, based on the results, there is no need to incorporate wetland-specific inputs when modelling similar landscape areal coverage.

3.5.2. Case study 2: 1A Creek Watershed

Watershed 1A differs from Willow Creek watershed in that 30% of the watershed area is occupied by wetlands. Over 75% of this area is occupied by a distinct wetland composition (the "toe seepage 2" map unit). The developed flow model for this watershed was good in simulating the average daily flow values (Figure 3-8), with R^2 values ranging from 0.81 to 0.99 for all the modeled data sets (Table 3-3). As before, the simulated daily flow values were used to model the *TP* concentration. The simulation was good in dry years, as indicated by 2002 and 2003 predictions (Figure 3-9). The peak locations were adequately replicated, providing no lag phenomena; however, the model overestimated *TP* concentration in the early spring of 2001 (a storm dominated year) and underestimated *TP* concentration in the summer of the same year.

It is likely that the large "toe seepage 2" area in the watershed can be hydrologically disconnected when the ground is still frozen. This in turn may delay water export from the wetland to the stream. It appears that the ANN model was able to capture this phenomenon, highlighting the robustness of the coupled time series/ANN approach when flow was being modeled. However, when modelling *TP* concentration, the model was not very successful in accurately simulating peak responses. This is likely because wetlands tend to accumulate sediment that has high phosphorus content over time. The sediment build-up continues until a high storm event washes it out to the nearest receiving stream. Thus, during early snowmelt, the wetland exports less sediment, being possibly disconnected (explaining model overestimation of the 2001 spring peak). Later in the season, when a storm washes accumulated sediment, phosphorus export is consequently higher (likely reflecting model underestimation of the 2001 summer peak).

3.6. Conclusions and Recommendations

An ANN stream flow model was devised and applied to two forested watersheds in Northern Alberta, Canada. The simulated daily flow values were then used to develop a predictive tool for the daily change in the *TP* concentration. The power of the developed models was verified by the high coefficient of multiple determination, the low root mean squared error, and the consistency in predicting the trends in data patterns for all the studied cases. The developed ANN flow models managed to successfully simulate average daily flow with R^2 exceeding 0.8 for all modeled data sets. ANN provided an adequate tool for modelling *TP* concentration attaining R^2 ranging from about 0.78 to 0.96 for all models. A three-slab hidden layer MLP ANN was designed and utilized in this study. It is believed that each slab can manipulate one of three distinct processes that control the system behavior (base flow, snowmelt, and rain events). This is an interesting conclusion, and is, thus, the topic of an ongoing study that examines the relation between different types of flow hydrographs and neural networks' architectures. The effect of *TP/Q* hysteresses was reasonably accounted for based on a hybrid spectral analysis/ ANN approach. The cross-correlation analysis was successful in highlighting the important time-lagged inputs despite the fact that the studied series were non-stationary.

The wetland area and composition appears not to influence average daily flow predictions, thus, there was no need to incorporate wetland-specific inputs for the study watersheds when daily flow was being modeled using ANN. However, results of *TP* concentration predictions for a watershed of high percentage of wetland areal coverage (1A Creek) suggested that watershed-specific inputs are needed to improve the ANN model predictions. Therefore, more information about the dynamics of phosphorus export from at least some wetland types is required for better representation of wetland characteristics in the development of *TP* ANN models.

Models like the ones developed in this study that use commonly available inputs, yet reasonably accurate, provide a useful tool for modelling ungauged watersheds. The concepts presented in this study can easily be extrapolated to other similar watersheds permitting flow and water quality predictions in response to climate change and landscape management practices. It can also offer a hydrologic link to the development of multi-objective forest management plans.

3.7. References

Band, L.E., Tague, C.L., Groffman, P., and Belt, K. 2001. Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export. Hydrol. Process. **15**: 2013-2028.

Brion, G.M., and Lingireddy, S. 2003. Artificial neural network modelling: a summary of successful applications relative to microbial water quality. Water Sci. Technol. **47**(3): 235-240.

Chanasyk, D.S., Whitson, I.R., Mapfumo, E., Burke, J.M., and Prepas, E.E. 2003. The impacts of forest harvest and wildfire on soils and hydrology in temperate forests: A baseline to develop hypotheses for the Boreal Plain. J. Environ. Eng. Sci. **2**: S51-S62.

Chatfield, C. 2001. Time Series Forecasting. Chapman and Hall CRC Press, New York, 267 pp.

Dingman, S.L. 2002. Physical Hydrology, second ed. Prentice Hall, Upper Saddle River, New Jersey, 646 pp.

El-Din, A.G., and Smith D.W. 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. **36**(5): 1115-1126.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation, second ed. Macmillian CollegePublishing, Inc., New York, 842 pp.

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus and phosphorus load from watershed characteristics using back-propagation neural networks. Acta Ecologica. **17**(1): 43-53.

Maier H.R. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and application. Environ. Modell. Softw. **15**: 101–124.

Maier, H.R., and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. **32**(4): 1013–1022.

Maier, H.R., Dandy, G.C., and Burch, M.D. 1998. Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in River Murray, South Australia. Ecol. Modell. **105**: 257-272.

Masters, T., 1995. Neural, novel and hybrid algorithms for time series prediction. John Wiley & Sons, Inc., 514 pp.

McEachern, P., Prepas, E.E., Gibson, J.J., and Dinsmore, W.P. 2000. Forest fire induced impacts on phosphorus, nitrogen, and chlorophyll a concentrations in boreal subarctic lakes Northern Alberta. Can. J. Fish. Aquat. Sci. **57**: 73-81.

Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural networks. Application of control and validation data series in Middle Loire River. Ecol. Modell. **120**: 141-156.

Nour, M.H., Smith, D.W., Gamal El-Din, M. and Prepas, E.E. 2006. Artificial neural networks and time series modelling of TP concentration in boreal streams: a comparative approach. J. Environ. Eng. Sci. **5**(S1): S39-S52.

Prepas, E.E., Burke, J.M., Whitson, I.R., Putz, G. and Smith, D.W. 2006. Associations between watershed characteristics, runoff, and stream water quality: hypothesis development for watershed disturbance experiments and modelling in the Forest Watershed and Riparian Disturbance (FORWARD) project. J. Environ. Eng. Sci. **5**(S1): S27-S37.

Prepas, E.E., Pinel-Alloul, B., Planes, D., Methot, G., Paquet S., and Reedyk, S. 2001. Forest harvest impacts on water quality and aquatic biota on the Boreal Plain: introduction to the TROLS lake program. Can. J. Fish. Aquat. Sci. **58**: 421-436.

Ryan, M., Muller, C., Di, H.J., and Cameron, K.C. 2004. The use of artificial neural networks (ANNs) to simulate N2O emissions from a temperate grassland ecosystem. Ecol. Modell. **175**: 189-194.

Shumway, R.H., and Stoffer, D.S. 2000. Time Series Analysis and its Applications. Springer, New York, 549 pp.

Ward Systems, 1996. NeuroShell 2 version 3, Ward Systems Group, Inc.,Frederick, Maryland, USA.

Wilson, H., and Recknagel, F. 2001. Towards a generic artificial neural network model fordynamic predictions of algal abundance in freshwater lakes. Ecol. Modell. **146**: 69-84.

Zealand, C.M., Burn D.H., and Simonovic S.P., 1999. Short term streamflow forecasting using artificial neural networks. J. Hydrol. **214**: 32–48.

Zhang, Q., Audrey, A.C., Shariff, R., and Stanley, S.J. 2004. Implementing artificial neural network models for real-time water color forecasting in a water treatment plant. J. Environ. Eng. Sci. **3**: S15-S23.

Zhang, Q., and Stanley, S.J. 1997. Forecasting raw water quality parameters for North Saskatchewan River by neural network modelling. Water Res. **31**: 2340-2350.

Table 3-1. Summary table for all models' inputs

Final Model	Inputs
Model 1 (*Q* for Willow)	$R_t, R_{t-1}, R_{t-2}, R_{t-3}, sin(2\pi vt), cos(2\pi vt), T_{max}, T_{mean}, T_{min}, dd_t, dd_{t-1}, dd_{t-2}, S_t, S_{t-1}, S_{t-2}$
Model 2 (*TP* for Willow)	$TP_{t-1}, sin(2\pi vt), cos(2\pi vt), T_{mean}, , \Delta Q_t, \Delta Q_{t-1}, \Delta Q_{t-3}$
Model 3 (*Q* for 1A)	$R_t, R_{t-1}, R_{t-2}, sin(2\pi vt), cos(2\pi vt), T_{max}, T_{min}, dd_t, dd_{t-1}, S_t, S_{t-1}$
Model 4 (*TP* for 1A)	$TP_{t-1}, sin(2\pi vt), cos(2\pi vt), T_{mean}, , \Delta Q_t, \Delta Q_{t-2}, \Delta Q_{t-3}, \Delta Q_{t-4}$

where: R_t, R_{t-1}, R_{t-2}, and R_{t-3} are the rainfall in mm at lags 0 through 3; T_{max}, T_{mean}, and T_{min} represents maximum, daily mean, and minimum air temperatures in °C, respectively; dd_t, dd_{t-1}, and dd_{t-2} are the cumulative degree days at lags zero to two; S_t, S_{t-1}, and S_{t-2} are the cumulative snowfall in mm for lags 0 through two; $\Delta Q_t = (Q_t - Q_{t-1})$, ΔQ_{t-1}, ΔQ_{t-2}, ΔQ_{t-3}, and ΔQ_{t-4} are the daily change in flow at lags 1, 2, 3 and 4, respectively.

Table 3-2. Summary table showing optimum ANN models' architecture and ANN internal parameters

	Model 1 (*Q* for Willow)	**Model 2 (*TP* for Willow)**	**Model 3 (*Q* for 1A)**	**Model 4 (*TP* for 1A)**
Scaling function	Linear, <<-1,1>>	Linear, <<-1,1>>	Linear, <<-1,1>>	Linear, <<-1,1>>
Optimum network (I-HG-HL-HGC-O)	15-4-4-4-1	8-5-5-5-1	11-5-2-5-1	7-7-5-7-1
Output activation function	tanh	Logistic	tanh	tanh
Training algorithm	BP	BP-BM	BP	BP-BM
Learning rate	0.2	Insensitive	0.15	Insensitive
Momentum coefficient	0.2	Insensitive	0.15	Insensitive

where: I denotes the input layer; HG, HL, and HGC are the Gaussian, logistic, and Gaussian complement slabs hidden layer, respectively; tanh is the hyperbolic tangent function; and << >> denotes an open interval.

Table 3-3. Statistical measures of models' performance

		Model 1 (*Q* for Willow)			Model 2 (*TP* for Willow)			Model 3 (*Q* for 1A)			Model 4 (*TP* for 1A)		
		T	**S1**	**S2**	**T**	**S1**	**S2**	**T**	**S1**	**S2**	**T**	**S1**	**S2**
S1 as testing data set	**R^2**	0.96	0.85	0.84	0.95	0.91	0.78	0.98	0.92	0.81	0.86	0.84	0.82
	RMSE	0.04	0.07	0.07	15	18	34	0.02	0.04	0.05	33	22	26
S2 as testing data set	**R^2**	0.94	0.84	0.8	0.96	0.81	0.79	0.99	0.81	0.82	0.81	0.82	0.78
	RMSE	0.05	0.06	0.08	14	31	27	0.02	0.05	0.06	37	36	24

T, training data set; S1, testing data set; S2, cross-validation data set; RMSE is in m^3/s for flow and in µg/L for *TP* concentration

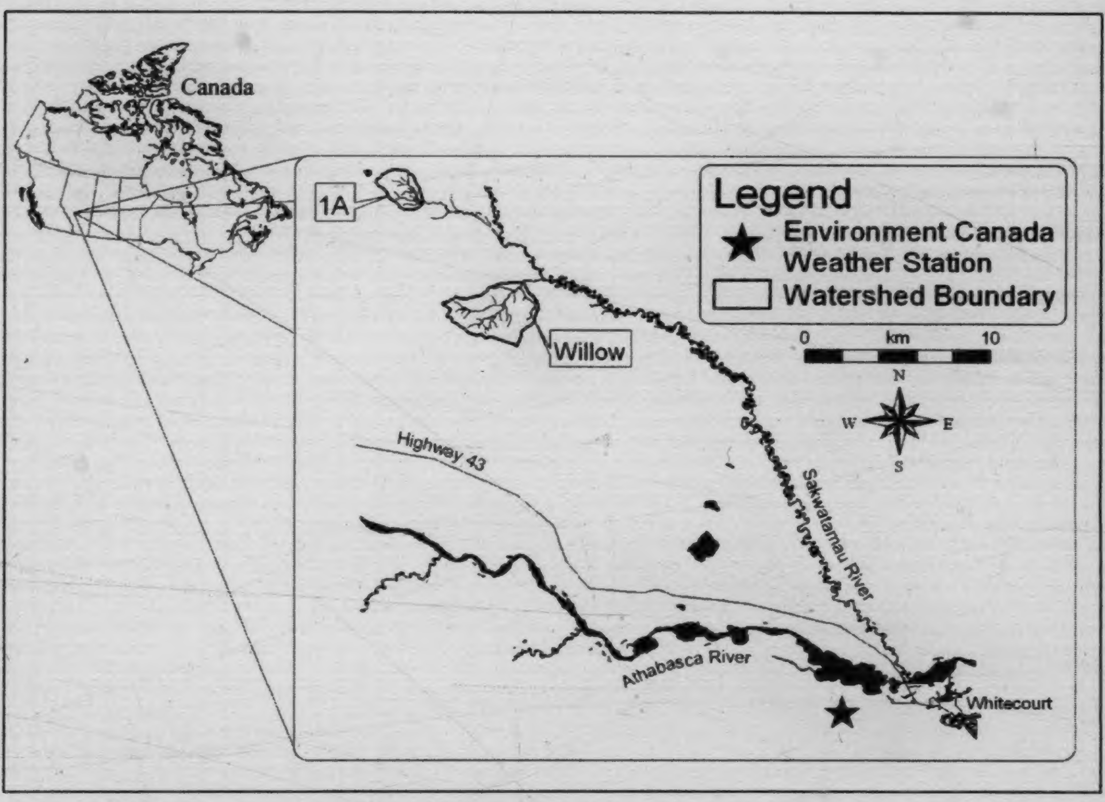

Canada
1A
Willow
Legend
Environment Canada Weather Station
Watershed Boundary
0
km
10
N
W
E
S
Highway 43
Sakwatamau River
Athabasca River
Whitecourt

Figure 3-1. The study area showing the two modeled watersheds

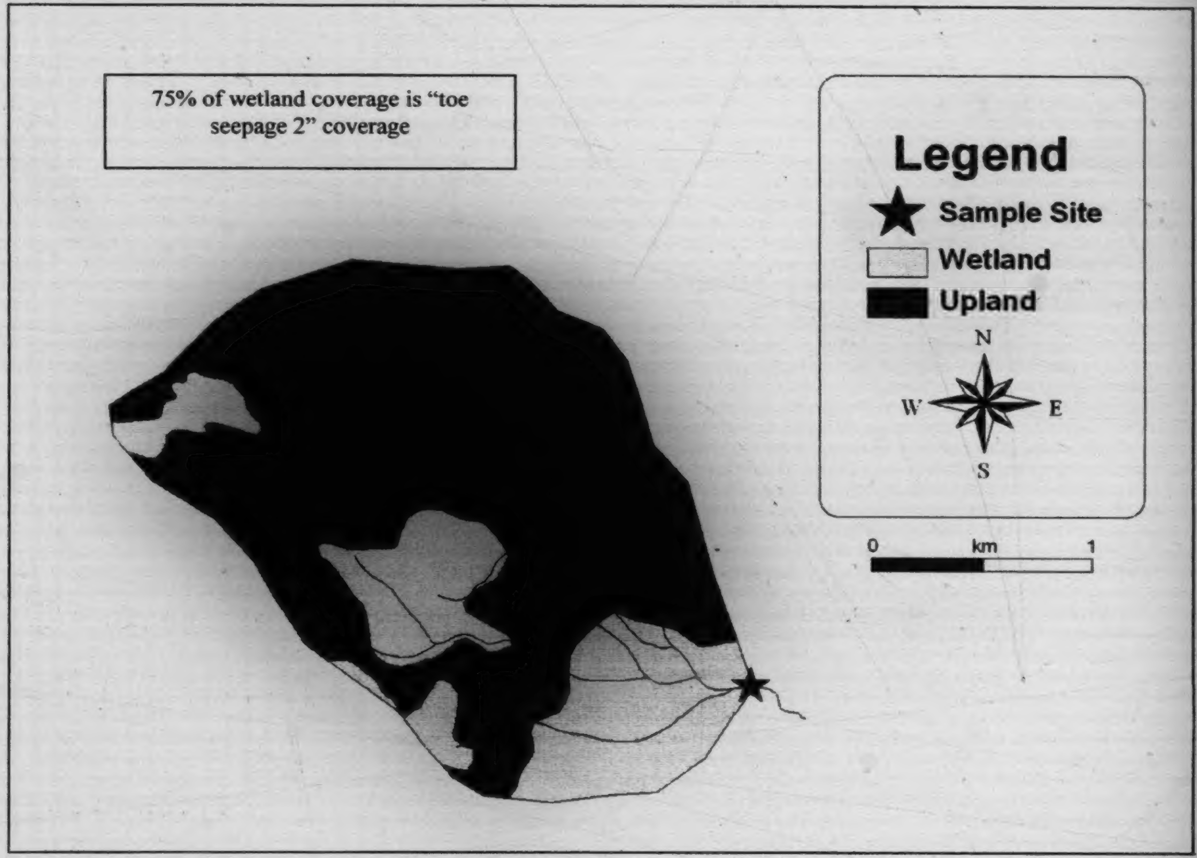

75% of wetland coverage is "toe seepage 2" coverage
Legend
Sample Site
Wetland
Upland
N
W
E
S
0
km
1

Figure 3-2. 1A Creek watershed soil map showing the wetlands locations within the watershed

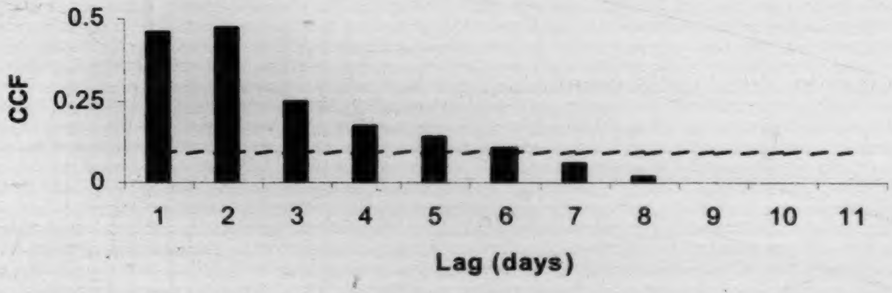


Figure 3-3. CCF of 1A Creek flow and *TP* concentration, dashed line is the 95% confidence boundary

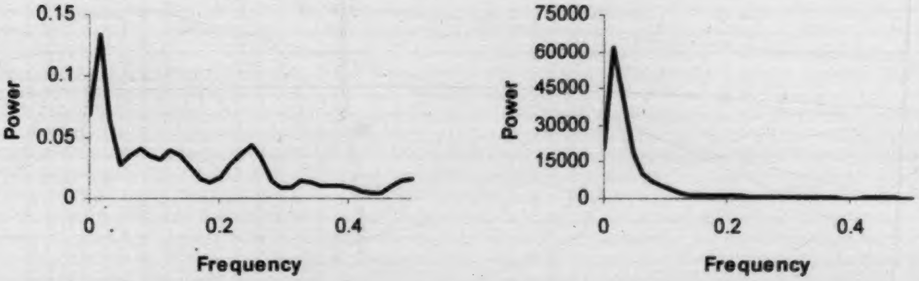


Figure 3-4. Power spectrum of 1A Creek flow in $[m^3/s]^2$ (left panel) and *TP* concentration in $[\mu g/l]^2$ (right panel)

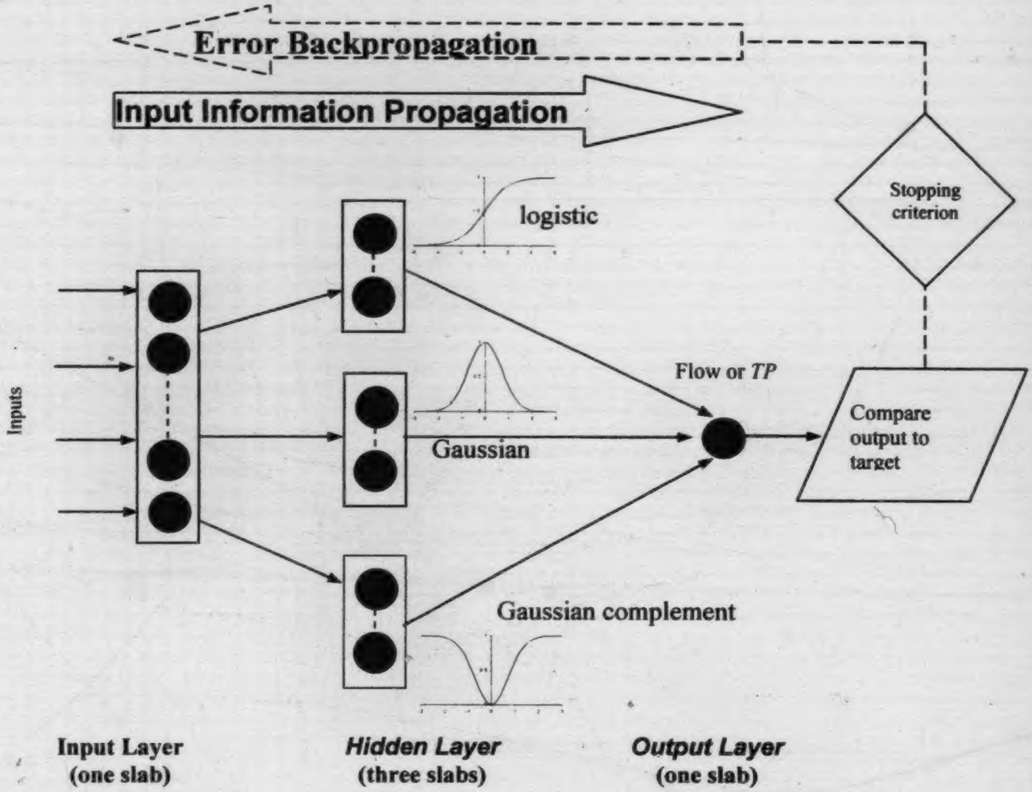

Error Backpropagation
Input Information Propagation
Stopping criterion
logistic
Inputs
Flow or TP
Compare output to target
Gaussian
Gaussian complement
Input Layer
(one slab)
Hidden Layer
(three slabs)
Output Layer
(one slab)

Figure 3-5. Schematic showing ANN optimum architecture for all four models

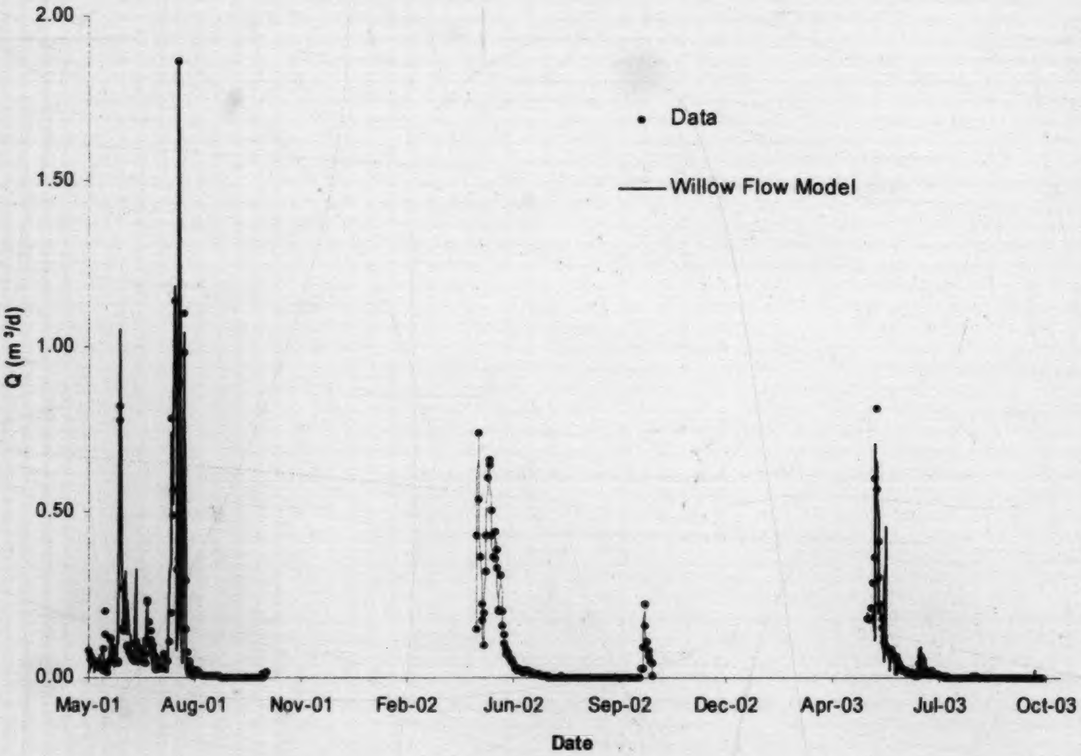


Figure 3-6. Measured versus ANN predicted flow hydrographs for the Willow Creek watershed

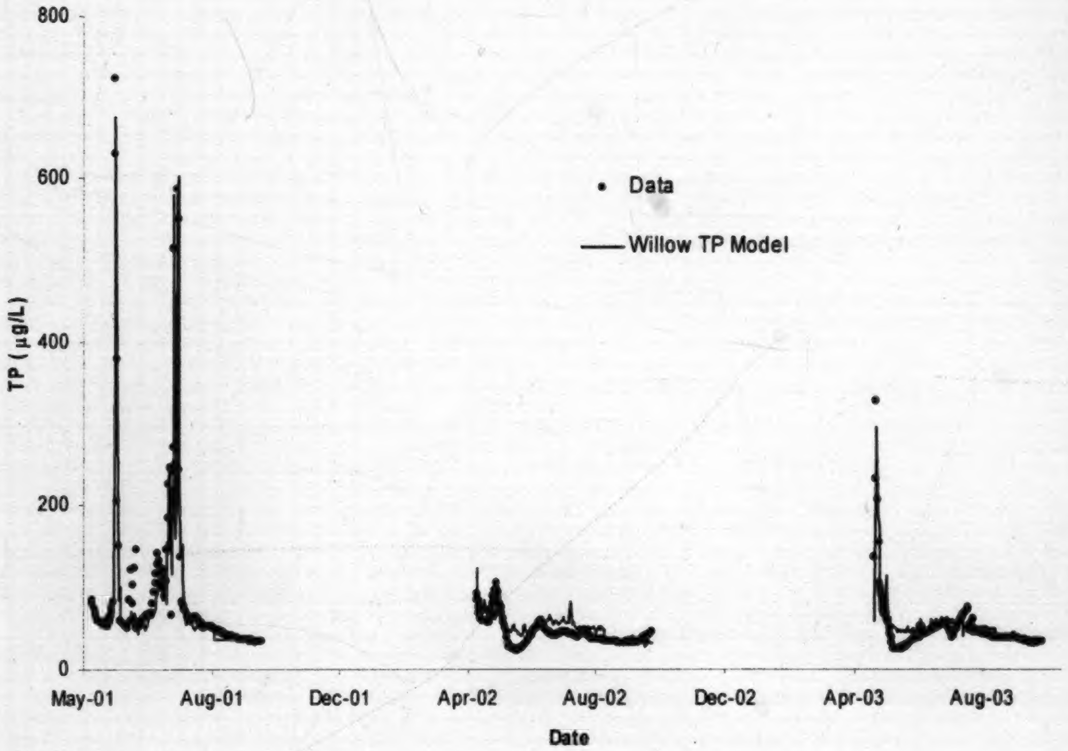


Figure 3-7. Measured versus ANN predicted *TP* concentration profile for the Willow Creek watershed

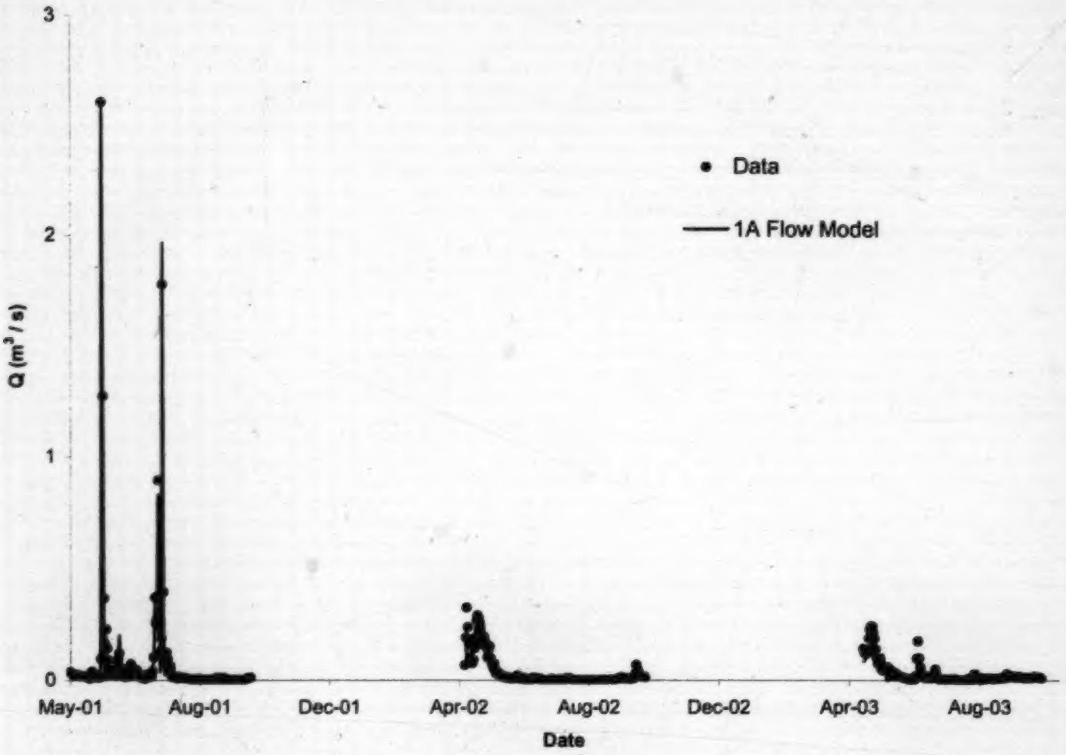


Figure 3-8. Measured versus ANN predicted flow hydrographs for the 1A Creek watershed

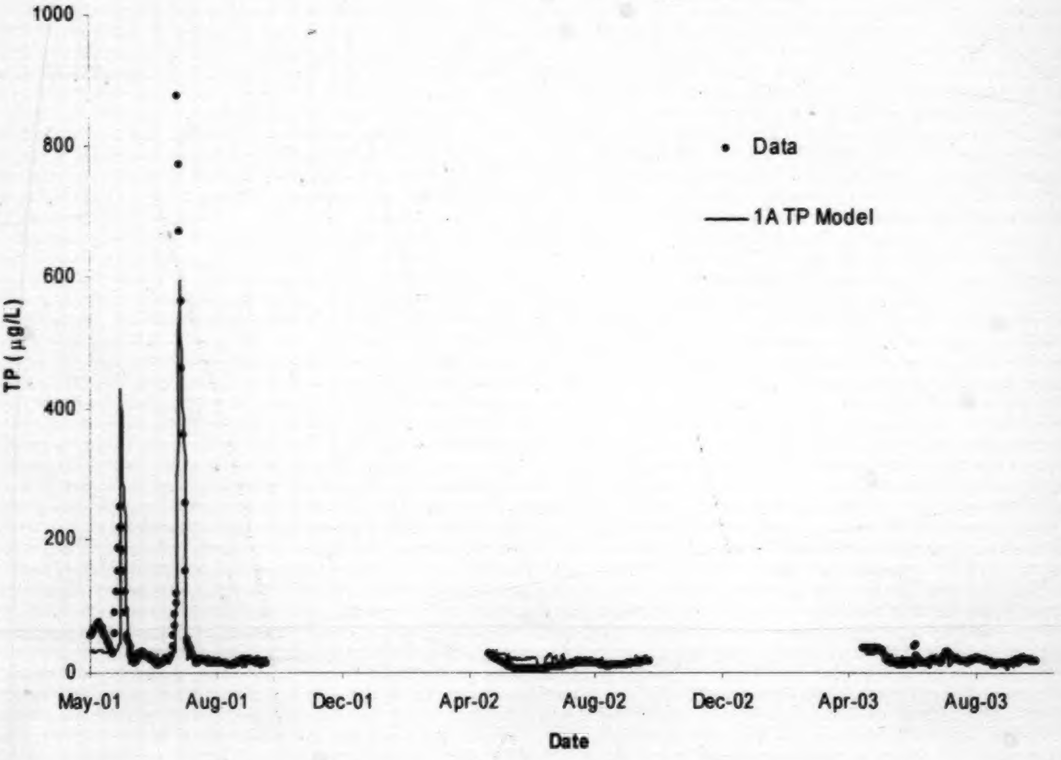


Figure 3-9. Measured versus ANN predicted *TP* concentration profile for the 1A Creek watershed

CHAPTER 4. GEOSTATISTICAL MAPPING OF PRECIPITATION: IMPLICATIONS FOR RAIN GAUGE NETWORK DESIGN

4.1. Introduction

The struggle to develop better models of hydrologic processes has been a continuing area of intense research for the last two decades. The precipitation time series, in particular, is an important input in hydrologic water quantity and quality models. However, networks measuring stations are typically sparse, and available data are insufficient to characterize the highly variable—in space and time—precipitation distribution, undermining the applicability of many modelling efforts (St-Hilaire et al. 2003; Singh and Woolhiser 2002; Tsintikidis et al. 2002). Therefore, reliable mapping of precipitation is crucial to the success of water quantity and quality modelling. Several sparse data interpolation techniques have been discussed in the literature. These interpolation methods include but are not limited to: The Thiessen polygon method (Thiessen 1911); inverse distance weight, IDW (Watson and Philip 1985); spline interpolation (Guenni 1997); and geostatistical or kriging techniques (Isaaks and Srivastava 1989; Journel 1989). However, geostatistical techniques are increasingly preferred because they allow one to capitalize on the spatial structure/correlation between neighbouring observations, to predict attribute values, and to quantify prediction uncertainty at unsampled locations (Diodato and Ceccarelli 2005; Apaydin et al. 2004; Goovaerts 2000 and 1999; Pardo-Iguzquiza 1998; Tabios and Salas 1985).

A version of this chapter has been published. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas E.E. 2006. Neural networks modelling of streamflow, phosphorous, and suspended solids: application to the Canadian Boreal forest. *Water Sci. Technol.* **53**(10): 91-99.

This paper compares five geostatistical interpolation techniques for mapping daily rainfall values: simple kriging (SK), ordinary kriging (OK), multi-Gaussian kriging (MGK), log-normal kriging (LNK), and kriging with an external drift (KED). The kriging methods were also compared to the more traditional IDW interpolation technique. All the interpolation schemes were compared in terms of cross-validation statistics. Sequential Gaussian simulation was then implemented to develop conditional stochastic simulation (100 realizations in a 250 x 250 m grid) of rain events. A multi-objective approach was finally introduced to suggest potential sites for the installation of further weather stations.

4.2. Study Area and Data Analysis

Rainfall information was acquired from 15 weather stations (Table 4-1) at close proximity to the forest watershed and riparian disturbance (FORWARD) project study area (see *J. Environ. Eng. Sci.* special issue Volume 2, 2003 for a detailed description of the project). The frequency of acquiring rainfall information varied among the monitored stations between one hour and one day. The time interval during which each station is operating within a year is highly variable. To overcome this problem, a code was implemented in Matlab to sum the sub-daily values and to present all rainfall data as daily representations. Rainfall data was available from May 2001 to October 2004. However, data was kept only if at least one of the rain gauges captured more than a trace value. A histogram of rainfall values revealed a highly skewed distribution (coefficient of skewness = 3.8, coefficient of Kurtosis = 19.7) with approximately 70% of the data close to the trace value. In addition to the precipitation information, a 30-m resolution digital elevation model (DEM) was acquired for the study area (Figure 4-1a). ESRI® ArcMap™ 9.0 was used to derive the slope from the DEM (Figure 4-1b).

4.3. Variography and Variogram Modelling

The discipline of Geostatistics is based on the random function (RF) concept (Deutsch 2002; Isaaks and Srivastava 1989; Journel 1989) whereby the set of unknown values is regarded as a set of spatially dependent random variables (RVs). Usually the RF definition is restricted to RVs related to the same attribute (rainfall in this case); hence, a second RF is needed to express the spatial variability of a second attribute (say, elevation). Geostatistical techniques replace the Euclidean distance ($h = \sqrt{x^2 + y^2}$) with a structural distance named the variogram ($2\gamma(h)$), which is specific to the attribute and the field under investigation. The variogram is a measure of spatial variability; it increases as samples become more dissimilar. The variogram and the covariance function ($C(h)$) are closely related (Equation 1) under the stationarity assumption (i.e., the mean and variance are assumed constant and independent of data location, and the covariance is assumed to be a function of the translation h and not the data locations).

[1] $C(h) = \sigma^2 - \gamma(h)$,

where σ^2 is the deemed stationary variance, $C(h)$ and $\gamma(h)$ are the covariance and the semivariogram calculated for *lag-h* , respectively. Although the covariance is what is needed for subsequent geostatistical modelling, it is a common practice to infer the semivariogram from data because it is easier to model. The covariance counterpart can then be calculated from Equation 1. Equation 2 demonstrates the relation used to estimate the semivariogram from sparse data:

$$[2] \qquad \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} [z(u) - z(u+h)]^2$$

where $N(h)$ is the number of data pairs separated by lag-h. In practice, some tolerance in h is identified to have enough data for constructing a reliable variogram. $z(u)$ and $z(u+h)$ are the rainfall intensities at locations u and $u+h$, respectively.

To identify the optimum lag spacing and lag tolerance, a histogram of the separation distances (h) between weather station data pairs was constructed (Figure 4-2a). It shows that the minimum h to be used is 10 km and that a reasonable one with respect to the number of data pairs is 35 km. The spatial structure/correlation of the precipitation is known to be anisotropic in most cases. Thus, a wind-rose type histogram (Figure 4-2b) and a colour-coded directional semivariogram (Figure 4-2c) were constructed and analyzed to assess the feasibility of including anisotropy in further modelling procedures. The wind rose shows that the data was fairly well distributed except for the direction of 120° from the East. The 30° from the East direction contained the largest number of data pairs. An ideal isotropic random function (RF) should show circular contours on the colour-coded semivariogram. According to the degree and direction of anisotropy, the circle will be distorted into an elliptical shape. Figure 4-2c shows a mild geometric anisotropy with a major axis in the direction of 80° from the East. Thus, based on these outcomes, directional variograms were constructed in the major and minor directions of anisotropy and were proven to be noisy due to insufficient data pairs in these directions. An omni-directional variogram was then constructed owing to the mild degree of anisotropy, and attempting to construct a well-suited variogram. Because the covariance must be known at all distances—not only at the separation distances between weather stations—a licit model that describes the experimental variogram is required. The sole constraint is that the variogram has to be modelled with a positive definite model in order to ensure the existence and uniqueness of the solution to the kriging set of equations. Figures 4-3a through 4-3c portray the experimental and modelled semivariograms for the original data, normal-score transformed data, and log-normal transformed data, respectively. The correlation structure of the data values is evidently similar in the three data spaces. Table 4-2 summarizes the parameters of the modelled variograms. All three models included a nugget effect, a short-scale exponential variogram model and a large-scale spherical model. However, the variance contributions and the range parameters were different in each case. The models were fitted to reach an asymptotic sill outside the data range, and thus should not be used for h > 135 km (maximum spacing supported by the data).

4.4. Kriging Techniques

Consider the problem of estimating the daily rainfall value ($Z(u)$) at an unsampled location u. The available information consists of same-day daily rainfall values at n locations (u_α, $\alpha = 1 .... n$). To create a map of the daily rainfall from measurements at sparse sampling stations, a least-square optimization technique (kriging) that can predict a random variable (RV) at an unsampled location with the objective of minimizing squared error was used. This section briefly discusses the different kriging algorithms used in this study. A more detailed discussion of the kriging methods can be found in many references (e.g., Deutsch 2002; Isaaks and Srivastava 1989; Journel 1989). We examined four univariate kriging techniques (SK, OK, MGC, LNK) and two multivariate kriging algorithms—KED using elevation as the external drift (KED-ELEV), and KED using slope as the external drift (KED-SLOPE). In all the kriging algorithms used, the rainfall spatiotemporal RF $\{Z(u),\ u \in study\ area\}$ was modelled as a collection of a finite number (*T*) of temporally correlated space RFs where *T* represents the number of days during which at least one of the rain gauges recorded more than the trace rainfall value within the duration of the study. Thus, spatial maps of the rainfall distribution were constructed only for the *T* time instances, and no time interpolation was possible without some additional modelling (Kyriakidis and Journel 1999; Kyriakidis et al. 2004).

4.4.1. Simple Kriging (SK)

The SK algorithm allows for the estimation of the daily rainfall depth ($Z^{*}(u)$ at an unsampled location *u* as a linear combination of neighbouring observations as expressed by Equation 3.

$$[3] \qquad Z^{*}_{SK}(u) = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(u) Z(u_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(u)\right] m ,$$

where $Z(u)$ is the RV model of daily rainfall at location *u*, the u_{α} values are the *n* data locations, *n* is the number of data points that went into kriging, $\lambda_{\alpha}(u)$, $\alpha = 1, \ldots, n$ are the SK weights used in estimating $Z(u)$ at location *u*, and *m* is the deemed stationary mean over the study area. The stationary mean was estimated from the sample data, and the SK weights were obtained by minimizing the error variance. This method is equivalent to solving a set of simultaneous equations known as the SK system of equations (Equation 4):

$$[4] \quad \sum_{\beta=1}^{n} \lambda_{\beta}^{SK}(u) C(u_{\beta} - u_{\alpha}) = C(u - u_{\alpha}), \alpha = 1, \ldots\ldots, n$$

where *C(h)* denotes the covariance function calculated at a lag-distance as reported between parentheses. The only information needed to formulate the kriging system is the covariance (*C(h)*). *C(h)* was calculated from the original-space variogram model (Table 4-2) followed by substitution into Equation 1. The strength of kriging methods is that one can get a sense of the prediction uncertainty by evaluating the location dependent error variance (σ_{SK}^2) as represented by Equation 5:

$$[5] \quad \sigma_{SK}^2(u) = \sigma^2 - \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK} C(u - u_{\alpha})$$

The above estimation procedure was repeated at each estimation node on a 250 x 250 m grid, and a daily precipitation map was produced.

4.4.2. Ordinary Kriging (OK)

The OK estimator is, in essence, a SK estimator in which the deemed stationary mean (*m*) is replaced by the location-dependent mean estimate $m^*(u)$. Thus, OK is already a nonstationary algorithm corresponding to a nonstationary RF with a variable mean but a stationary covariance. The same procedures implemented in the SK case were repeated here by replacing Equations 3, 4, and 5 with their OK counterparts represented by Equations 6, 7, and 8, respectively.

[6] $$Z^{*}_{OK}(u) = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{OK}(u) Z(u_{\alpha})$$

[7] $$\begin{cases} \sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(u) C(u_{\beta} - u_{\alpha}) + \mu(u) = C(u - u_{\alpha}), \alpha = 1, \ldots\ldots, n \\ \sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(u) = 1 \end{cases}$$

where *$\mu(u)$* is a Lagrange multiplier added to account for the additional constraint $\sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(u) = 1$.

[8] $$\sigma^{2}_{OK}(u) = \sigma^{2} + \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_{\alpha}^{OK}(u) \lambda_{\beta}^{OK}(u) C(u_{\beta} - u_{\alpha}) - 2\sum_{\alpha=1}^{n} \lambda_{\alpha}^{OK}(u) C(u - u_{\alpha})$$

4.4.3. Multi-Gaussian Kriging (MGK)

Prediction accuracy is typically better if the sample histogram does not suffer high skewness; thus, a normal-score transformation was used to transfer the sample histogram to a standard Gaussian histogram by matching the p-quantiles of the two distributions (details can be found in Deutsch (2002)). SK was then performed on the transformed values. The results were then back-transferred to the original data space. However, normal-score transformation is reversible only if no spikes or ties (a large proportion of the data holding a constant value) are present in either the original or the target histograms (Deutsch 2002). Our data indicated about 70% close to zero value, a large spike. Thus, two despiking algorithms were tried prior to applying the normal-score transform. The first was random despiking implemented by using the public domain GSLIB software (Deutsch and Journel 1998). The second was a local average despiking algorithm that computes local averages within local neighbourhoods centred at each tied data value. The data was then de-spiked according to local averages; tighed values in high valued areas would then get larger ranks than those in low-valued areas. The implementation of the MGK was similar to that of SK, except for dealing with the normal transferred data values and normal space semivariogram ($\gamma_2(h)$) instead of the original ones.

4.4.4. Log-Normal Kriging (LNK)

The histogram of the original daily rainfall values clearly demonstrated a log normal distribution. Therefore, carrying out a log-normal transform prior to kriging was worthwhile. The natural logarithm of the data was calculated, and then SK (Equations 3 to 5) was conducted on the transformed data. The results of SK on the transformed data were then back-transferred to the original space. The log-normal space ($\gamma_3(h)$)—rather than the original data space ($\gamma_1(h)$)—was used to solve the kriging system.

4.4.5. Kriging with an External Drift (KED)

Precipitation tends to increase by increasing elevation and may vary in response to the variation of the slope and aspect. Hence, incorporating a secondary variable that identifies such variation may enhance rainfall estimation. KED is a simple and efficient kriging algorithm for including a secondary variable in the estimation of the primary variable. If the secondary variable is known to vary smoothly in space over the study area, and if this variable is known at all locations of interest (u_α), the algorithm can be applied and can produce fairly robust results. Analogous to other kriging algorithms, the spatial estimates of rainfall was computed as a linear combination of the surrounding data. However, two conditions were added to the optimization of weight values to assure unbiasedness and utilization of the secondary RF. Equations 9 to 11 represent the estimation process, the kriging system, and the kriging variance in the case of using KED, respectively:

[9] $$Z^{*}_{KED}(u)=\sum_{\alpha=1}^{n}\lambda_{\alpha}^{KED}(u)Z(u_\alpha)$$

[10] $$\begin{cases}\sum_{\beta=1}^{n}\lambda_{\beta}^{KED}(u)C(u_\beta-u_\alpha)+\mu_o(u)+\mu_1(u)y(u_\alpha)=C(u-u_\alpha),\alpha=1,......,n\\ \sum_{\beta=1}^{n}\lambda_{\beta}^{KED}(u)=1\\ \sum_{\beta=1}^{n}\lambda_{\beta}^{KED}(u)y(u_\beta)=y(u)\end{cases}$$

[11] $$\sigma^2_{KED}(u)=\sigma^2+\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_{\alpha}^{KED}(u)\lambda_{\beta}^{KED}(u)C(u_\beta-u_\alpha)-2\sum_{\alpha=1}^{n}\lambda_{\alpha}^{KED}(u)C(u-u_\alpha),$$

where $y(u)$ represents the secondary variable (in this study, either elevation or slope), and $\mu_o(u)$and $\mu_1(u)$ are the Lagrange multipliers that account for conditional optimization. KED does not require data transformation, and thus the original space semivariogram was utilized in this case. In this case, there is no need to calculate the cross-variogram between the primary and the secondary variables because such variogram plays no role in the kriging algorithm. Only an estimate of the secondary variable at all estimation nodes is required.

4.5. Inverse Distance Weights (IDW) Interpolation

IDW interpolation is a widely used interpolation technique, which assumes that objects close to one another are more alike than those that are far apart. Thus, IDW interpolation presumes that each measured location has a local influence that diminishes with distance. To predict a rainfall value for any unmeasured location, IDW interpolation uses the surrounding weather stations' measurements; however, the stations closest to the prediction location will have the greatest influence on the predicted values. At each time step, the code calculates the separation distance between the point of interest and each of the relevant stations by using Equation 12. The code then calculates the corresponding weights (*w_i*) by using Equation 13, and finally an estimate of rainfall at the point of interest is made by using Equation 14. Estimating a measure of prediction uncertainty, akin to the kriging variance, is not possible when the IDW scheme is used as an interpolation technique.

[12] $$r_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

[13] $$w_i = \frac{1/r_i^2}{\sum_{j=1}^{n} \frac{1}{r_j^2}}$$

[14] $$\hat{R}_{(x,y)}(t) = \sum_{i=1}^{n} w_i \cdot R_i(t),$$

where (x,y) are the coordinates of the point where rainfall is to be estimated, $\hat{R}_{(x,y)}(t)$ is the estimated average daily rainfall in mm/d at time t, n is the number of weather stations to be used in estimation, i is the weather station number, and $R_i(t)$ is the measured daily rainfall at station i and time t in mm/d.

4.6. Evaluation of The Different Interpolation Methods

Daily maps were produced according to the seven aforementioned algorithms. Because assessing the performance of each algorithm by visually examining the produced maps is subjective, the performance of the kriging algorithms was alternatively assessed by using cross-validation statistics. The idea is to temporarily remove one weather station at a time from the data set and then re-estimate the removed value from the remaining data using the alternative interpolation algorithms. Pearson's correlation coefficient (*r*) and the root mean squared error (RMSE) were computed for each algorithm and were used to compare the different interpolation algorithms. Table 4-3 summarizes the statistical performance of the cross-validation data in response to each interpolation scheme. In our case study, incorporating information from DEM into the rainfall estimation did not enhance the rainfall predictions. This result can be explained by the relatively mild slope of the study area. The correlation coefficients of rainfall/elevation and rainfall/slope time series were very small for the study domain ($<$ 0.2) explaining why a secondary variable did not enhance the results. The results also indicated that SK, OK, LNK, and IDW are comparative algorithms with respect to prediction performance; however, OK produced slightly better results in terms of *r*, and IDW produced the best results in terms of RMSE and came next to OK with regards to *r*. MGC was the worst univariate estimator, mainly due to large data spikes. Although data despiking was carefully done, the Gaussian back-transfer was disrupted due to the high proportion of data spikes.

Despite the smoothness of the developed maps, it is important to observe the spread of the rainfall values. Figure 4-4 depicts two kriging maps constructed by using OK for June 28, 2002 and July 8, 2004. The map shows that for June 28, 2002, the SW region received more rainfall than the NE whereas for July 8, 2004, the NW region experienced more intense rainfall than the SE. Interestingly, the prevailing wind directions for these days were NE and SE (arrows shown on Figure 4-4), likely leading to this distribution of rainfall. Apparently, the modelled variogram managed to grasp this feature of the data without including any wind information in the modelling process. It is critical to note that the calculated kriging variance is a function of the data configuration and not the data values (notice the similarity of the two maps of kriging variance in Figure 4-5), and, therefore, this variance can be used to provide implications about the expected error surface in response to different weather network configurations.

4.7. Sequential Gaussian Simulation (SGS)

The produced kriging maps confirmed that kriging, when used as a mapping algorithm, acts as a low-pass filter that tends to smooth out the details and extreme values of the original data set. Thus, the actual values of a RF are more random than their corresponding kriged estimates. Stochastic simulation, on the other hand, can correct the missing variance and can produce a number of equiprobable realizations of the rainfall while honouring data values at the measurement locations and assuring the reproduction of the data joint spatial continuity. In the current study, we utilized the SGS algorithm (details in Deutsch 2002) to produce 100 equiprobable maps of rainfall and then calculated the probability of exceeding a certain threshold of rainfall over the study domain. Examples of such probability maps are displayed in Figure 4-6 for thresholds of 50 and 5 mm/d. As expected, the proportion of the landscape with a high probability of receiving above 5 mm/d of rainfall is much larger than the proportion with a high probability of exceeding 50 mm/d.

4.8. Future Rain Gauge Network Design

In Alberta, rain gauge network design is experience-based. Decisions are based mainly on the proximity to specific locations of interest and the accessibility to the proposed sites, rather than on a methodological scientific approach. On the other hand, geostatistitians typically utilize the kriging variance as the sole criterion for selecting extra sampling sites. This approach is incomplete because the operational needs are also important for an engineering decision. In this study, we combined the two approaches to identify the optimal locations for new rain gauges. The approach is based on overlaying the map of the kriging variance, the DEM, and land use/land cover and road networking maps in a Geographical information system (GIS) framework. Zones of high kriging variance were delineated first, then checked for accessibility by using the DEM, the land use/land cover, and the road networking maps. The common areas of favourable features were proposed as future sites.

The FORWARD project models flow and water quality for 16 watersheds in the

Canadian Boreal forest. Four weather stations were recently installed in the area as part of the project. Identification of future sites for the installation of further weather stations is required. The abovementioned approach was utilized for this case study. The four FORWARD weather stations were added to the 15 stations used in this study. The resulting map of kriging variance was obtained by using OK (Figure 4-7a). It is evident that the variance is close to zero at close proximity to the weather stations and increases with the distance from each weather station. The NW and SE regions of the FORWARD study area were identified as zones of higher kriging variance. The map of kriging variance was then overlaid on the DEM map (Figure 4-7b), and the road network map (Figure 4-7c) and the areas of high kriging variance, good accessibility, and as close as possible to the FORWARD study watersheds were identified and displayed as black rectangles in Figure 4-7c. Since other measurements of interest (like those of air temperature and solar radiation) take place in each weather station, no attempt should be made to optimize for the number and exact locations of future stations until the same procedure has been conducted for the other parameters of interest as well. Upon completion for all parameters of interest, an optimization algorithm that targets the minimization of a cost function which reduces cost and kriging variance by choosing optimal locations of new stations can be conducted for areas that are commonly favourable for all the parameters of interest.

4.9. Conclusions and Recommendations

Kriging is now commonly used as a mapping technique; however, practitioners are often confused by all the available kriging methods and other interpolation schemes.

The choice of the optimal interpolation algorithm should be guided primarily by the characterization of the data under study. Adequate understanding of the theory and intrinsic assumptions of each technique is, therefore, critical. We examined four univariate kriging techniques (SK, OK, MGC, LNK) and two multivariate kriging algorithms—KED using elevation as the external drift (KED-ELEV), and KED using slope as the external drift (KED-SLOPE), as well as the traditional IDW interpolation scheme for the estimation of daily rainfall in a 250 m x 250 m grid over a 750 Km^2 area in the Canadian Boreal forest. The results indicated that multivariate kriging did not enhance daily rainfall estimation skill. This finding can be explained by the relatively mild slope of the study area. SK, OK, LNK, and IDW were proven to be comparative algorithms with respect to prediction performance; however, OK produced slightly better results in terms of Pearson's correlation coefficient. IDW outperformed OK in terms of RMSE but came next to it in performance with regards to *r*. The strength of OK, as compared to IDW, was in the ability to estimate a measure of prediction uncertainty by evaluating the kriging variance. However, for our case study, if a prediction error estimate is not required, IDW can be used as an interpolation technique without jeopardizing accuracy. MGC was the worst univariate estimator, likely due to the high percentage of data spikes. Although data despiking was carefully done, the Gaussian back-transfer was disrupted due to the high proportion of spikes. SGS was then implemented to produce 100 equiprobable maps of rainfall, and the probability of exceeding nominal thresholds of rainfall over the study domain was calculated. Such information can be used as inputs to hydrologic and water quality models to address the uncertainty in the modelled parameter in response to the uncertainty in rainfall information. A multi-objective approach, based on overlaying the map of the kriging variance, the DEM, and land use/land cover and road networking maps in a GIS framework to identify the areas of commonly favourable features, was proposed to identify potential future sampling locations. The approach was applied to the FORWARD study area, and favourable regions for the installation of further weather stations were identified.

4.10. References

Apaydin, H., Sonmez, F.K., and Yildirim, Y.E. 2004. Spatial interpolation techniques for climate data in the GAP region in Turkey. Clim. Res. **28**(1): 31-40.

Deutsch C.V. 2002. *Geostatistical reservoir modelling.* Oxford University Press, New York.

Deutsch C.V., and Journel A.G. 1998. *GSLIB – Geostatistical Software Library and User's Guide*. 2nd edn., Oxford University Press, New York.

Diodato, N. and Ceccarelli, M. 2005. Interpolation processes using multivariate geostatistics for mapping of climatological precipitation mean in the Sannio Mountains (southern Italy). Earth Surf. Process. Landf. **30**(3): 259-268.

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **228**(1-2): 113-129.

Goovaerts, P. 1999. Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena. **34**(3-4): 227-242.

Guenni, L. 1997. Spatial interpolation of stochastic weather model parameters. J.

Environ. Manage. **49**(1): 31-42.

Isaaks, E., and Srivastava, S. 1989. An introduction to applied geostatistics. Oxford Univerity Press.

Journel, A.G. 1989. Fundamentals of geostatistics in five lessons. Volume 8 Short Course in Geology, American Geophysical Union.

Kyriakidis, P.C., and Journel, A.G. 1999. Geostatistical space-time models: A review. Math. Geol. **31**(6): 651-684.

Kyriakidis, P.C., Miller, N.L., and Kim, J. 2004. A spatial time series framework for simulating daily precipitation at regional scales. J. Hydrol. **297**(1-4): 236-255.

Pardo-Iguzquiza, E. 1998. Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. Int. J. Climatol. **18**(9): 1031-1047.

Singh, V.P., and Woolhiser, D.A. 2002. Mathematical modeling of watershed hydrology. J. Hydrol. Eng., **7**(4): 270-292.

St-Hilaire A., Ouarda T.B.M.J., Lachance M., Bobee B., Gaudet J., and Gignac C. 2003. Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrol. Process.,* **17**: 3561-3580.

Tabios G.Q. and Salas, J.D. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour. Bull. **21**(3), 365-380.

Thiessen, A.H. 1911. Precipitation for large areas. Mon. Wea. Rev. **39**: 1082-1084.

Tsintikidis, D., Georgakakos, K.R., Sperfslage, J.A., Smith, D.E., and Carpenter, T.M.

2002. Precipitation uncertainty and raingauge network design within Folsom Lake watershed. J. Hydrol. Eng., 7(2): 175-184.

Watson, D.F., and Philip, G.M. 1985. A Refinement of Inverse Distance Weighted Interpolation. Geo-Processing. 2(4): 315-327.

Table 4-1. Summary table of weather stations' locations

Station ID	Station Name	Latitude	Longitude	Elevation (m)
EA	Eagle	54.4569N	116.4397W	1048
FA	Ft Assiniboine	54.3414N	114.8083W	640
GM	Goose Mountain	54.7506N	116.0332W	1385
IM	Imperial	54.4666N	115.571W	1202
MB	Mayberne	53.8613N	116.6645W	1453
PS	Pass Creek	54.2289N	116.8388W	1081
SD	Swan Dive	54.7283N	115.3545W	1241
SO	Shining Bank	53.8064N	115.9389W	899
W1	Freeman Auto	54.5561N	115.2972W	821
W3	Meekwap Auto	54.6249N	116.6597W	836
W4	Windfall Auto	54.1883N	116.2497W	808
W5	Fox Creek Auto	54.3974N	116.8025W	850
WC	Whitecourt	54.0324N	115.7197W	1172
ZU	Whitecourt	54.0836N	115.7856W	741
ENVCAN	Environment Canada	54.15N	115.78W	782.4

Table 4-2. Summary of variogram models utilized in this study

Space	Variogram model
Original space	$\gamma_1(h)=\begin{cases}0.15+0.23\left\{1-\exp\left(-\frac{3h}{45000}\right)\right\}+0.62\left\{1.5\frac{h}{280000}-0.5\left(\frac{h}{280000}\right)^3\right\}, & \text{if h} \leq 280000\\ 1.00, \ \text{if h} \geq 280000\end{cases}$
Normal space	$\gamma_2(h)=\begin{cases}0.12+0.13\left\{1-\exp\left(-\frac{3h}{15000}\right)\right\}+0.75\left\{1.5\frac{h}{320000}-0.5\left(\frac{h}{320000}\right)^3\right\}, & \text{if h} \leq 320000\\ 1.00, \ \text{if h} \geq 320000\end{cases}$
Log-normal space	$\gamma_3(h)=\begin{cases}0.17+0.12\left\{1-\exp\left(-\frac{3h}{17000}\right)\right\}+0.71\left\{1.5\frac{h}{350000}-0.5\left(\frac{h}{350000}\right)^3\right\}, & \text{if h} \leq 350000\\ 1.00, \ \text{if h} \geq 350000\end{cases}$

Table 4-3. Statistical performance of utilized kriging algorithms

	IDW	SK	OK	MGC	LNK	KED-ELEV	KED-SLOPE
Correlation coefficient	0.74	0.74	0.76	0.67	0.75	0.73	0.65
RMSE (mm)	3.04	4.90	4.78	5.5	4.81	4.96	5.74

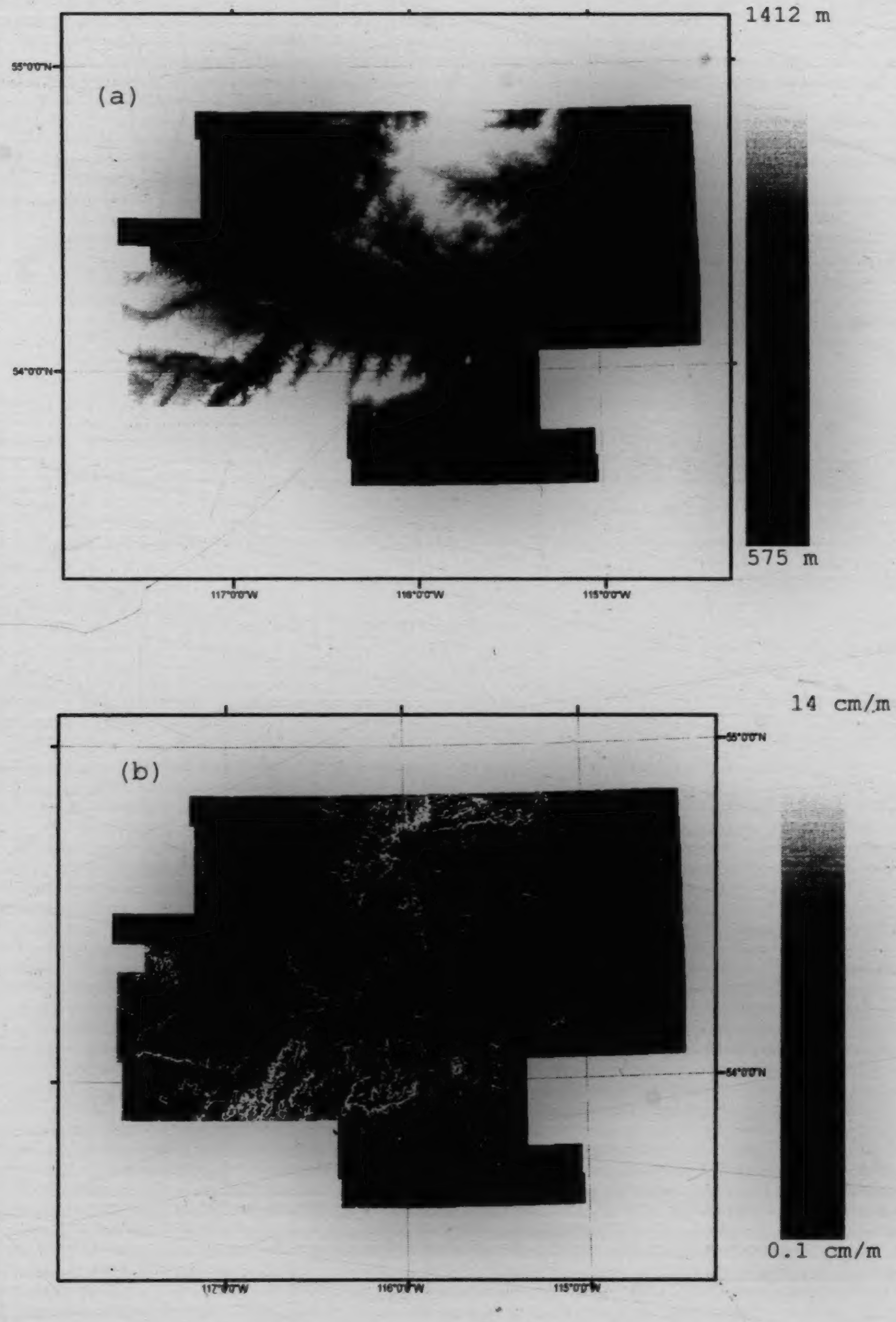

1412 m
55°0'0"N
(a)
54°0'0"N
575 m
117°0'0"W
116°0'0"W
115°0'0"W
14 cm/m
55°0'0"N
(b)
54°0'0"N
0.1 cm/m
117°0'0"W
116°0'0"W
115°0'0"W

Figure 4-1. (a) 30-m resolution DEM and (b) slope distribution in the study area

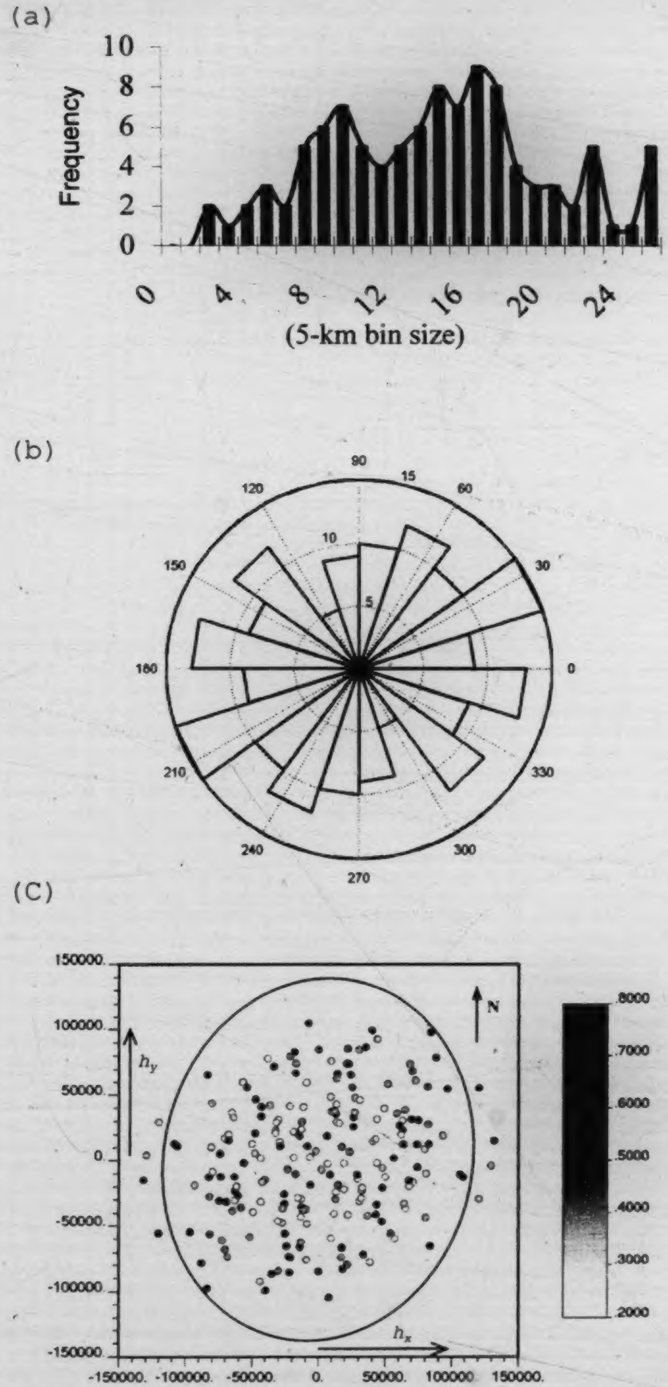


Figure 4-2. (a) Histogram of weather station pairs of separation distances, (b) wind-rose type histogram of separation distances, and (c) colour coded directional semivariogram

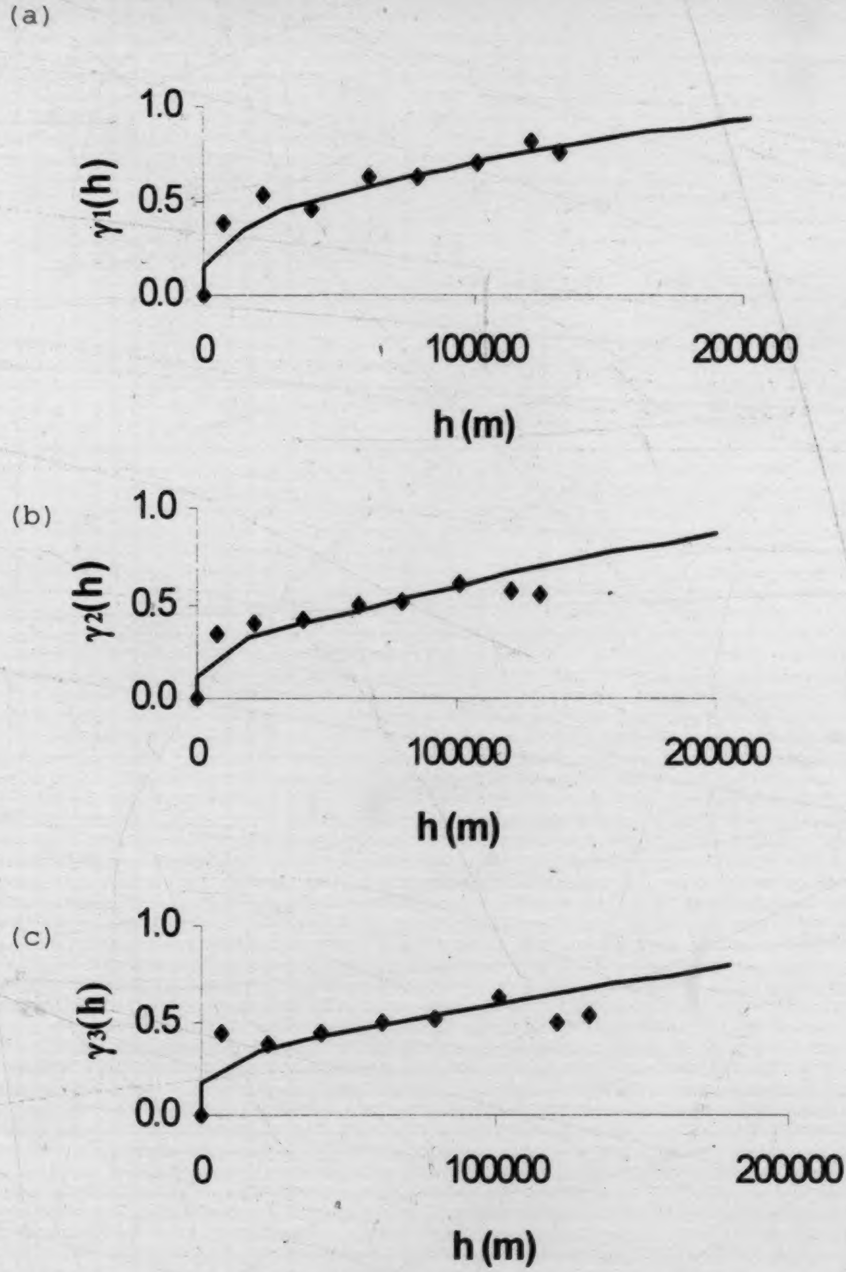


Figure 4-3. Experimental and modelled omni-directional semivariograms for (a) original data space, (b) normal-score transformed data, and (c) log-normal transformed data; solid lines represent modelled semivariograms

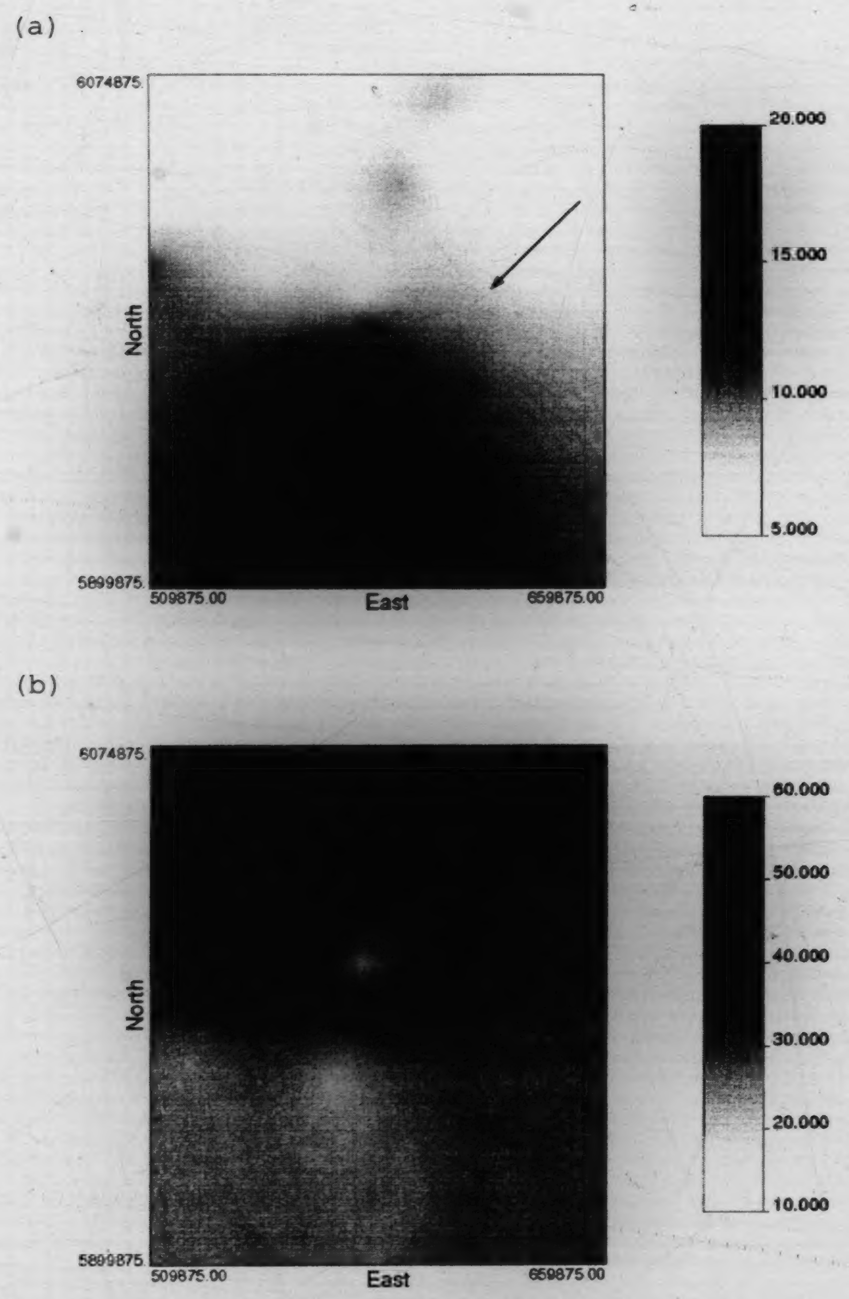

(a)
6074875
5899875
North
509875.00
East
659875.00
20.000
15.000
10.000
5.000
(b)
6074875
5899875
North
509875.00
East
659875.00
60.000
50.000
40.000
30.000
20.000
10.000

Figure 4-4. Kriging maps of rainfall in mm obtained by OK for (a) June 28, 2002 and (b) July 8, 2004; arrows showing prevailing wind direction

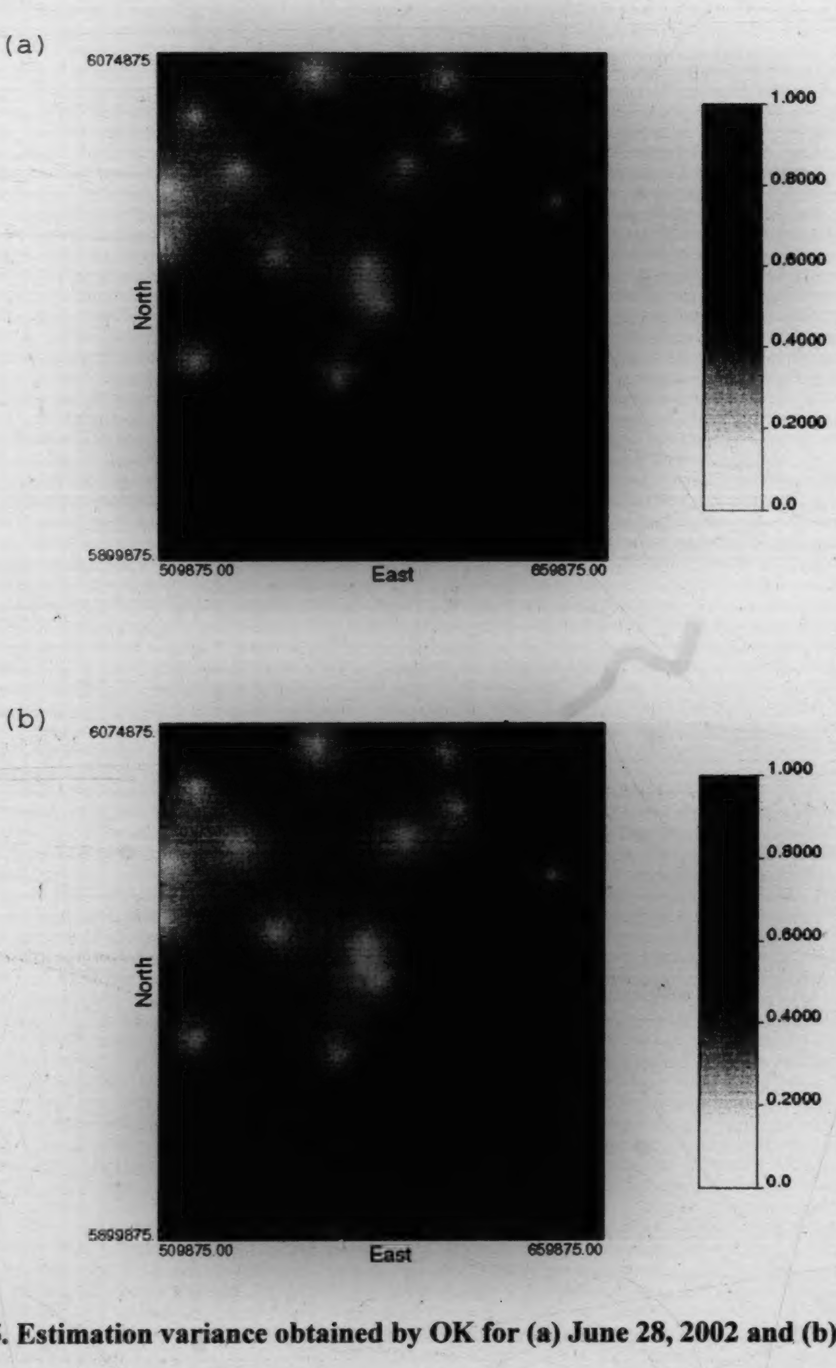

(a)
6074875
5899875
North
East
509875.00
659875.00
1.000
0.8000
0.6000
0.4000
0.2000
0.0
(b)
6074875
5899875
North
East
509875.00
659875.00
1.000
0.8000
0.6000
0.4000
0.2000
0.0

Figure 4-5. Estimation variance obtained by OK for (a) June 28, 2002 and (b) July 8, 2004

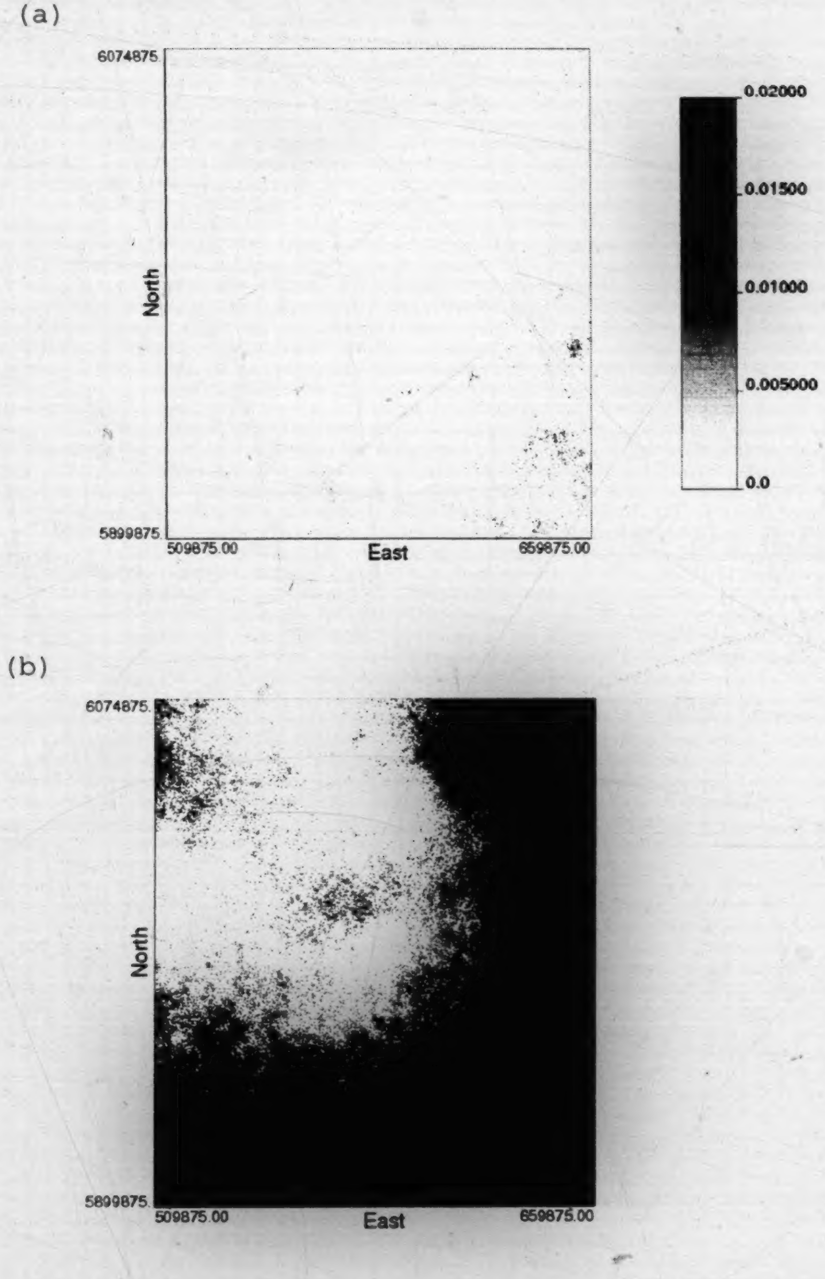

(a)
6074875.
North
5899875.
509875.00
East
659875.00
0.02000
0.01500
0.01000
0.005000
0.0
(b)
6074875.
North
5899875.
509875.00
East
659875.00

Figure 4-6. Spatial probability distribution of exceeding (a) 50 mm/d and (b) 5 mm/d

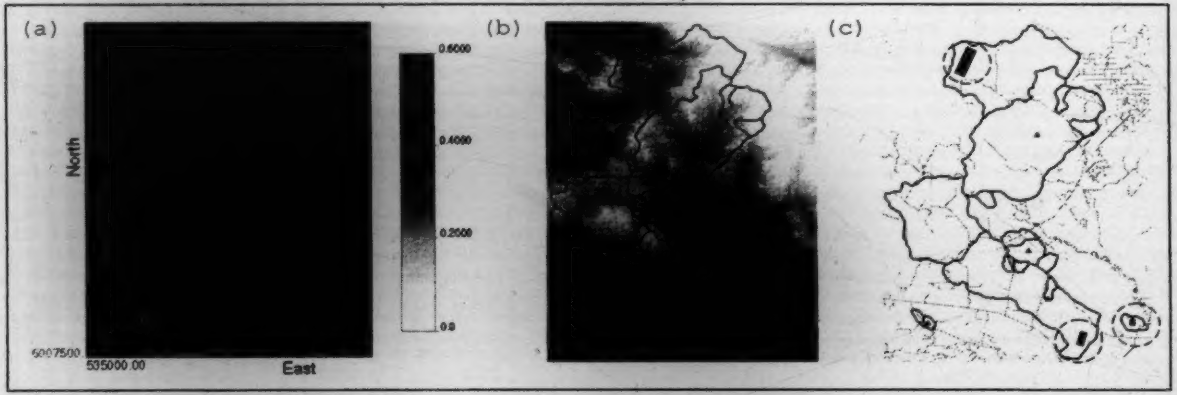

(a)
North
East
5007500
535000.00
0.6000
0.4000
0.2000
0.0
(b)
(c)

Figure 4-7. (a) Kriging variance distribution after adding the four FORWARD weather stations (b) DEM, and (c) proposed new sites for future weather stations (black rectangles); ▲ denoting current FORWARD study weather stations

CHAPTER 5. TOWARDS A GENERIC ARTIFICIAL NEURAL NETWORK MODEL FOR DYNAMIC PREDICTIONS OF DAILY STREAMFLOW IN UNGAUGED WATERSHEDS: INTRODUCING A NEW MEASURE OF HYDROLOGIC SIMILARITY

5.1. Introduction

The application of artificial neural networks (ANNs) in hydrological modelling has been the topic of over 300 refereed publications in the last two decades. The ASCE task committee (2000a and 2000b) and Maier and Dandy (2000) published comprehensive reviews of pertinent work prior to the late 1990s. Most researchers used either a feed-forward multi layer perceptron (FF-MLP) ANN or a recurrent neural network (RNN) for modelling daily streamflow. However, the feed-forward MLP trained with the error backpropagation (BP) algorithm was by far the most widely used network architecture and training algorithm (Castellano-Mendez et al. 2004; Anctill and Rat 2005; Riad et al. 2004; Agarwal and Singh 2004; Riad and Mania 2004; Baratti et al. 2003; Kisi 2004; Tokar and Markus 2000). In all reviewed cases, a sliding window of rainfall and recently observed flow values were utilized to forecast future flow values. Tawfik (2003) constructed a FF-MLP ANN model to predict the River Nile's flow utilizing information from four gauging stations along the River Nile. She used the past values for the flow at the four stations to predict the flow at Aswan, Egypt.

A version of this chapter has been submitted for publication. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006. Towards a generic artificial neural network model for dynamic predictions of daily streamflow in ungauged watersheds: introducing a new measure of hydrologic similarity. *J. Environ.Eng. Sci.* (submitted 03/2007).

Sudheer et al. (2003) demonstrated that the performance of a MLP-ANN could significantly be improved by applying an appropriate data transformation to the historic time series prior to model formulation. The authors were successful in simulating peak flows; however, their algorithm's need for antecedent flow information limits its use in ungauged watersheds. Anctil et al. (2004) proposed the use of soil moisture index as an auxiliary ANN input to the typically used rainfall and flow inputs attempting to account for the low-frequency hydrologic processes. Including such information enhanced their models' prediction ability, specifically in dry weather periods for the Leaf River in the USA and the Serein River in France.

Kumar et al. (2004) compared a RNN to a feed-forward MLP for single step ahead and multiple step ahead streamflow forecasts of an Indian River catchment and concluded that the RNN outperformed the FF-MLP network. Chang et al. (2002) and Cahng et al. (2004) developed a recurrent neural network model for flow prediction one hour and two hours in advance. They compared their models' prediction ability to that of time series models and concluded that the ANN models outperformed the time series class of models.

Owing to the complexity of hydrological processes, and motivated by the ability of artificial neural networks (ANNs) to model complicated non-linear relationships, all previous efforts demonstrated that ANN models performed at least comparatively to, if not better than, other deterministic and statistical models. However, most of these studies, albeit successful in simulating streamflow and forecasting flow at different lead times, failed to address the topic of modelling the streamflow of ungauged watersheds. The reviewed modelling efforts used past values of flow to predict the future ones. This class of models, although important for the real-time forecasting of gauged watersheds, cannot provide flow predictions in ungauged watersheds due to the lack of pertinent inputs.

In Canada and elsewhere, the prediction of daily streamflow is important for evaluating downstream hydrologic impacts, simulating the impact of extreme floods and droughts, and thus for safeguarding against any expected adverse consequences. Providing the resources to gauge all watersheds of interest is practically impossible; thus, a class of models that could simulate the response of ungauged watersheds with reasonable accuracy is important for effective watershed management and planning. Hence, the objectives of this study were (1) to develop a neural network modelling algorithm capable of modelling ungauged watersheds, (2) to apply the developed model to four watersheds in the Canadian Boreal forest, (3) to give an example of the applicability of the approach for an ungauged watershed case study, and (4) to find a reasonable indicator of hydrologic similarity that can guide model transferability. Initial results from this work was presented in the eighth international conference on the application of artificial intelligence to civil, structural and environmental Engineering (Nour at al. 2005).

5.2. Research Area and Database

The study area is located in the Virginia Hills, Alberta, Canada (Figure 5-1). As part of the forest watershed and riparian disturbance (FORWARD) project, the monitoring of the daily streamflow of four small watersheds (with a basin area of 5 to 130 km^2) began in the year 2001 and is still in operation. The four studied basins are 1A, 5.1 km^2;

Cassidy, 5.9 km^2; Willow, 15.6 km^2, and Two Creek, 129.4 km^2. The province of Alberta is covered by over 300 weather stations (mostly from fire towers and Environment Canada meteorological stations). Rainfall information was acquired from 15 weather stations at close proximity to the study area (Figure 5-1). The frequency of acquiring the rainfall varied among the monitored stations between 1 hr to 1 day. Also, the time interval during which each station was operating within a year was highly variable. To overcome this problem, a code was written to add up sub-daily values and to present all rainfall data as a daily representation.

5.3. Artificial Neural Networks Model Development

A parsimonious ANN model was systematically constructed in three phases: (1) data pre-processing, (2) model construction, and (3) model evaluation. The first phase aims at identifying the input variables, exploring all input/output data patterns, highlighting the data's main trends and features, and pinpointing any irregularities that would require further investigation. In the second phase, the modeller should carefully select the optimum ANN topology, the best training algorithm, should optimize the network internal parameters and data division among training (calibration), testing (validation), and cross-validation data sets. Finally, the third phase requires statistical and graphical assessment of the candidate model performance.

5.3.1. Data Pre-processing and Input Determination

In general, streamflow is highly correlated with time, exhibits an annual and seasonal cyclic variation, and passes through hystereses loops. Thus, when modelling streamflow, the inputs should reflect causality, time correlation, and the seasonal periodicity. Adequate understanding of the physical processes driving the modelled parameter is critical for identifying the causal inputs. Time series analysis (both the time domain and the frequency domain) can adequately identify time-lagged inputs and can feed the ANN model with information to reflect hystereses loops and the seasonal cyclic nature.

5.3.1.1 Causal inputs

The objective behind the choice of cause/effect type inputs used in this study was to identify a surrogate for each component of the water cycle. However, to be able to construct an efficient modelling tool for ungauged watersheds, all inputs should be easily acquired at a reasonable cost. Streamflow is dictated mainly by the processes responsible for the catchment's fast response (as manifested by peaks of the flow hydrograph), namely, rainfall and snowfall; and also by lower-frequency information (e.g.: soil moisture fluctuation, subsurface flow, and actual evapotranspiration) that can allow for the mapping of much slower processes associated with exchanges at the soil and vegetation surfaces. Constrained by the readily available data for the Canadian ungauged watersheds, the authors selected rainfall (R) and snowfall to reflect the catchement's fast response regimes, and selected the air temperature (T) as a surrogate for the solar energy available for evapotranspiration and snowmelt.

The local variations of rainfall can be significant even for a small area. Variations of up to 50% between recorded rainfalls at a given time were detected in the study area. Thus, in the devised models, the inverse distance weighted (IDW) interpolation scheme was utilized to better represent the rainfall at each modelled catchment.

Inverse distance weighted (IDW) interpolation is a widely used interpolation technique. It assumes that objects close to one another are more alike than those that are far apart. Thus, IDW interpolation presumes that each measured location has a local influence that diminishes with distance. To predict a rainfall value for any unmeasured location, IDW interpolation uses the surrounding weather stations' measurements; however, the stations closest to the prediction location will have the greatest influence on the predicted values. Figure 5-2 shows a hypothetical area with *n* weather stations and a location (x,y) where *R* is to be predicted by using data from the surrounding stations.

Eqs. 1 through 3 summarize the approach. A procedure was adopted to detect the *n* closest weather stations to the centroid of each of the four modelled watersheds at each time step, and thus accommodating the inconsistency in the time period of operation of each of the surrounding stations. At each time step, the code calculates the separation distance between the point of interest and each of the relevant stations by using Eq. 1. The code then calculates the corresponding weights (w_i) by Eq. 2, and finally an estimate of rainfall at the point of interest is made by Eq. 3:

$$[1] \quad r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

$$[2] \quad w_i = \frac{1/r_i^2}{\sum_{j=1}^{n} \frac{1}{r_j^2}}$$

$$[3] \quad \hat{R}_{(x,y)}(t) = \sum_{i=1}^{n} w_i \cdot R_i(t),$$

where (x,y) are the coordinates of the point where rainfall is to be estimated, $\hat{R}_{(x,y)}(t)$ is the estimated average daily rainfall in mm/d at time t, n is the number of weather stations to be used in estimation, i is the weather station number, and $R_i(t)$ is the measured daily rainfall at station i and time t in mm/d.

This approach produced a time series of daily rainfall for each of the modelled watersheds. The temperature (daily maximum, T_{max}; daily minimum, T_{min}; and daily average, T_{avg}) and the snowfall were available from a nearby Environment Canada weather station (Figure 5-1).

At high latitudes, snowmelt plays a key role in streamflow, especially in early spring. Thus, snowmelt has to be explicitly accounted for when constructing the streamflow model. Snowmelt is typically estimated by using either the energy balance approach or the temperature-index approach. However, due to the difficulty and the expenses of fulfilling the former's data requirements, the latter approach is the most extensively used in the literature (Dingman 2002). The temperature-index approach estimates snowmelt as a linear function of the average air temperature. The logic relies on the strong correlation between the solar radiation and the corresponding air temperature during snowmelt. This technique equates the daily snowmelt (ΔM) as a linear function of the mean air temperature provided that this temperature exceeds a base temperature for that day (Eq. 4). Therefore, during a specified time interval (*t*), the snowmelt is a function of the total of the degree-days (dd_t) summed up for the days when the temperature was above a baseline temperature (typically taken as zero), as represented by Eq. 5.

$$[4] \quad \Delta M = \begin{cases} f(T_{avg} - T_b), & T_{avg} \geq T_b \\ 0, & T_{avg} < T_b \end{cases}$$

$$[5] \quad dd_t = \sum_{i=0}^{i=L-1} (T_{avg(i)} - T_{b(i)}).(t_{i+1} - t_i),$$

where ΔM is the daily snowmelt in m, T_{avg} is the daily average air temperature in °C, T_b is a base temperature typically taken as 0 °C, L is the number of days during which $T_{avg} \geq T_b$, dd_t are the total degree days at time t in °C.day, and $(t_{i+1} - t_i)$ is typically taken as 1 day.

The cumulative snowfall at time t (S_t) was used to reflect the available snow depth, representing the amount of snow available for melting. The cumulative degree-days (dd_t) were used to provide an integrated measure of the heat energy available for snow melting. Thus, the cumulative snowfall and the degree-days can act as surrogates for the temperature-index snowmelt approach and are therefore used as inputs for the developed flow models:

5.3.1.2 Time-lagged inputs

In an atempt to evaluate the strength of the relation between the flow and the potential time-lagged input variables, a cross-correlation analysis was performed. The time-lagged variables of the causal inputs that are correlated to the flow were used as additional model inputs to represent the highly correlated nature of the studied time series. A statistical cross-correlation function (CCF) was calculated for each flow time series and the corresponding causal inputs (Eq. 6).

$$[6] \quad CCF = \frac{\sum_{t=1}^{n-h}(u_{t+h}-\bar{u})(v_t-\bar{v})}{\sqrt{\left[\sum_{t=1}^{n}(u_t-\bar{u})^2\right]\left[\sum_{t=1}^{n}(v_t-\bar{v})^2\right]}},$$

where n is the number of data points; h is the time lag in increment of days; t is the time expressed in integer intervals of days; u_t and u_{t+h} denote observations of variable u at times t and $t+h$, respectively; v_t and v_{t+h} represent observations of variable v at times t and $t+h$, respectively; and $\bar{u}$ and $\bar{v}$ are the mean of both u and v, respectively.

The time-lagged inputs with a CCF value higher than the 95% significant CCF were taken into consideration in model building. Figure 5-3 summarizes the CCF (*Q*, *R*) for the four studied watersheds. An interpretation of the CCF can be represented as it is in Figure 5-3b, where *Q(t)* is correlated to *R(t)* and *R(t-1)* for the Cassidy watershed. This correlation means that, for Cassidy, not only will its causal inputs be used in model building, but *R(t-1)* will be added to the vector of inputs to account for the time series behavior. A similar interpretation for the other watersheds can be made from Figure 5-3. However, due to the non-stationary nature of the studied series, the used CCF is not very accurate, and thus the time-lagged inputs were used as guidelines, and the possibility of eliminating the less correlated inputs was further investigated when building a candidate ANN model.

5.3.1.3 Inputs reflecting flow periodicity and Q/R hystereses loops

It was observed that the studied flow series exhibited an annual and sub-annual cyclic nature. Also the data exemplified the hystereses loops typical of recorded flow and rainfall. Additional inputs, to represent this seasonal behaviour, must be incorporated into the construction of a streamflow ANN model in order not to feed the model with contradictory information. (For instance, the yield from a given rainfall event on an initially dry catchment is lower than that from the same rainfall event on catchments which have recently received precipitation.).

Spectral analysis is used to efficiently manipulate a cyclic time series. The variance profile over the frequency, usually referred to as the "power spectrum", was constructed to identify the frequency that contributed most to the variance. This frequency (ν) was used later to help the model to dynamically change the input/output function according to the seasonal variation. Eqs. 7 to 9 demonstrate the calculation procedure of the power spectrum of a given flow time series:

$$[7] \quad \hat{f}(\upsilon) = \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} [X_c^2(\upsilon_k + l/n) + X_s^2(\upsilon_k + l/n)]$$

$$[8] \quad X_c(\upsilon_k + l/n) = n^{-1/2} \sum_{t=1}^{n} Q_t \cos\left(2\pi(\upsilon_k + l/n)t\right)$$

$$[9] \quad X_s(\upsilon_k + l/n) = n^{-1/2} \sum_{t=1}^{n} Q_t \sin\left(2\pi(\upsilon_k + l/n)t\right),$$

where n is the number of data points, t is the time in integer days, Q_t is the daily flow observation at times t, ν_k is the frequency of interest in units of d^{-1}, L is a smoothing parameter that should be assigned an odd number and should be fairly small relative to n, and X_C and X_S are the cosine and sine Fourier transforms, respectively.

Figure 5-4 depicts the flow's power spectrum for the four watersheds. It can be concluded that the frequency that contributed most to the power spectrum was in the range of 0.01 to 0.04 d^{-1}, corresponding to a period of 25 to 100 days. A value of 0.033 d^{-1}, corresponding to the month-to-month variation within the typical annual cycle, was used in order not to lose necessary information. Two additional ANN model inputs, $\sin(2\pi\frac{t}{12})$ and $\cos(2\pi\frac{t}{12})$, were included to account for the cyclic nature of the studied series. Through the sign of the two inputs, the model is expected to be able to identify the season under study (for example, a positive value for both the sine and cosine inputs identifies the winter season), and with the aid of the magnitude of these parameters, the model is believed to identify the month within each season as illustrated by Figure 5-5.

5.3.2. Model Construction

The aim of this phase is to build a parsimonious model. Such a model should not be limited to mapping the data used during its development but should also be able to generalize by mapping other data sets. Thus, the choice of the training algorithm, the network architecture, and the model internal parameters (number of hidden layers, number of neurons, type of scaling and activation functions, learning and momentum rates, and stopping criterion), and the division of data into training (calibration), testing (validation), and cross-validation data sets are crucial to achieve a robust model. A thorough description of the model's construction phase is presented in Nour et al. (2006b) and Maier and Dandy (2000).

5.3.2.1 Data division

According to the size of the available data, they must be split into different sets for the testing and validation of the ANN model. Without doubt, all data sets should be representative of the same population. When using an ANN, the available data are usually divided into two data sets for testing and validating the network (Maier and Dandy 2000). However, based on our experience and provided that sufficient data are available, the authors recommend operating in a cross-validation mode by dividing the data into three sets in the ratio of 3:1:1 for training, testing, and cross-validating the model, respectively. The training data set is used to calibrate the model by updating the neural network weights. The testing set is used to help in selecting a robust model out of the developed candidate models by determining when to stop training. Finally, the cross-validation data set is used to test the generalization ability of the chosen model through its application to a third data set. In this study, the models developed for the four watersheds were based on dividing the data in the ratio of 3:1:1 for training, testing, and cross-validating each model. The split was based on an algorithm that targets a similar frequency distribution of each data set, with any extreme and rare values being assigned to the training data set.

5.3.2.2 Identifying optimum network architecture

Maier and Dandy (2000) and Gamal El-Din and Smith (2002) provided a systematic approach for identifying the network architecture. This approach, with some modifications, was adopted in this study. Earlier work has found that a typical feed-forward (FF) multi-layer perceptron (MLP) ANN with a single hidden layer that utilizes one activation function in its processing elements (nodes) cannot accurately map the streamflow in higher latitudes (Nour at al 2006b). The complex nature of the streamflow system in higher latitudes is due mainly to the dynamic change of the driving forces of flow. At times, rainfall is the main driving force, but at other times, either the base flow or the snowmelt may control the flow values. Thus, a modification was made by applying a FF-MLP ANN with one hidden layer with processing elements that use more than one activation function. The approach proposed in Maier and Dandy (2000) and Gamal El-Din and Smith (2002) was then used to optimize the number of nodes and their corresponding activation functions.

5.3.2.3 Training algorithm

Two training algorithms were tested in this study: (1) the typical gradient descent BP algorithm that uses a learning rate and a momentum coefficient to control the training speed and to facilitate moving towards a global minimum in the error surface

(Haykin 1994), and (2) a BP algorithm with a batch update (BP-BU) technique. The NeuroShell 2 software package was used to train the models (Ward Systems Group 1996). In the batch mode of BP learning, training proceeds through an entire epoch (i.e. it cycles through all of the training data set patterns) before the weights are updated. The advantage of the second algorithm (the BP-BU algorithm) is that it is insensitive to both the learning rate and the momentum coefficient, giving flexibility to less experienced modelers (Gamal El-Din and Smith 2002).

5.3.2.4 Stopping criterion

In BP learning, minimizing an error function, mean squared error (MSE) in this case, is desirable; however, this process is complicated due to a typically multi-local minima error surface. Attempting to build a robust model that does not memorize the training data, we used the testing data set statistical performance measured by the MSE to dictate when to stop training. Typically, the repeated training iterations successively enhance the network's performance in the training data set, but the testing data set performance has an optimal point beyond which the statistical performance deteriorates again. Figure 5-6 demonstrates the approach. Training continued as long as the error of the testing data set was continuously decreasing, and was halted when this error started to increase even if that of the training data set was still decreasing, but the connection weights were always adjusted based on the training data set.

5.3.3. Model Evaluation

In this phase, the developed model's performance was evaluated by using statistical and graphical means. There is no single measure of "goodness-of-fit" statistic that can adequately describe model performance, and thus, an array of complementary measures was used for a complete assessment of model performance (Legates and McCabe Jr. 1999; Fox 1981; Willmot 1981). In this study, several relative- and absolute-error measures have been incorporated: percent relative bias, RB(%) (Eq. 10); the root mean squared error, RMSE (Eq. 11); the mean absolute error, MAE (Eq. 12); the square of Pearson's correlation coefficient, r^2; the coefficient of multiple determination, R^2 (usually referred to as the coefficient of efficiency (Nash and Sutcliffe 1970) as denoted by Eq. 13; and the second-order index of agreement, d_2 (Willmott et al. 1985) represented by Eq. 14:

$$[10] \quad RB(\%) = 100\left(\frac{\sum_{i=1}^{N} P_i - \sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} O_i}\right)$$

$$[11] \quad RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2}$$

$$[12] \quad MAE = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - P_i\right|$$

$$[13]\quad R^2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$$

$$[14]\quad d_2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(|P_i - \overline{O}| + |O_i - \overline{O}|)^2},$$

where P_i and O_i are the predicted and the measured streamflow values at time i, respectively; and $\overline{O}$ is the mean of the measured streamflow for the entire time period.

All six measures of "goodness-of-fit" were calculated for the training, the testing, and the cross-validation data sets and assessed to judge the candidate model performance. A graphical representation of the measured and the predicted flow hydrographs were then investigated to highlight zones of poor performance. In addition, the possible reasons for the poor simulations of some data regions were identified, and model improvements were attempted.

5.4. Model Application

The proposed algorithm was applied to model the daily streamflow of the four study watersheds. Causal inputs, time-lagged inputs, and inputs reflecting flow periodicity and *Q*/*R* hystereses loops were identified for each watershed. Table5-1 summarizes the inputs used for each of the modelled watersheds.

By following the guidelines provided in the ANN model development section, a robust feed-forward MLP ANN model was developed for each of the studied watersheds. Successful model training was achieved for all the cases by using a back-propagation training algorithm with batch update. A single hidden layer with three activation functions produced the best results for all the modelled watersheds, likely because each of the three distinct flow regimes (base flow, storm events, and snowmelt) is best manipulated with a separate activation function. Table 5-2 presents the architecture of the best model for each watershed.

Six statistical measures of "goodness-of-fit" were used to evaluate the performance of the devised models (Table 5-3). All models provided relatively low bias in the order of 10% with the exception of the cross-validation data sets of 1A watershed (-18%). Both RMSE and MAE were small as compared to peak flow reflecting the high performance of the devised models. The index of agreement, d_2, and the coefficient of multiple determination, R^2, exceeded 0.92 and 0.71, respectively for all studied data sets. These comparatively high values for all training, testing, and cross-validation data sets for all candidate models reflect the superiority of the modelling approach and its good generalization ability.

Figures 5-7 through 5-10 show the modelled versus the measured flow hydrographs for the four studied watersheds. For all watersheds, the modelled and measured hydrographs were in good match. The smaller the watershed, the more accurate the prediction was. The flow patterns (at peak locations) were accurately simulated for all devised models, with summer events being better represented by the models than spring snowmelt events (2004 Two Creek is an exception, as the spring runoff was better simulated than the summer runoff.). Although these models were constructed from low-cost readily available inputs, all models performed fairly well, providing a successful algorithm for modelling the streamflow based on meteorological information and efficient time series manipulation of this information without the need to include past flow values in the model vector of inputs.

Many times, when conducting a data-driven modelling approach, the parameters that are physically important might turn out to be rather trivial in the actual modelling application. Thus, it is important to assess the relative importance of model inputs with respect to their influence on the model output to make sure that the devised models are consistent with our conceptual understanding of the modelled system. In this study, the "weights" method as described by Garson (1991) was used to identify the relative contribution of inputs on the modelled daily streamflow. Rainfall, *R*, (summation of all time-lagged rainfall inputs) was found to be the main contributor for modelling daily streamflow for the Willow, Two Creek, and 1A watersheds (Figure 5-11). The Cassidy watershed was an exception where snowmelt inputs (degree-day, *dd*; snowfall accumulation, S(ct); and temperature information)) were the main contributors. The Cassidy streamflow time series, unlike the other three watersheds, spanned the period from 2002 to 2004 (all relatively dry years as compared to 2001). In such dry periods, the hydrograph is merely derived by base flow and snow melting explaining the relatively high importance of these inputs in case of the Cassidy watershed. The inputs reflecting *Q*/*R* hystereses were important in all four models and being more important in the Cassidy watershed as displayed in Figure 5-11.

5.5. Application to an Ungauged Watershed

The proposed algorithm managed to adequately simulate the streamflow of four watersheds relying on easily accessed information, highlighting the possibility of using such models in modelling ungauged watersheds. The application of this class of models requires delineating watershed basins by using GIS and digital elevation models (DEM). The generated basins should then be grouped together based on watershed characteristics and hydrologic similarity. A sample of each category should be gauged to formulate a representative model. Each representative model can then be used in a predictive mode to simulate the hydrologic impacts in all mapped ungauged watersheds with similar characteristics.

To test this approach, the model initially developed for the 1A watershed was used in a predictive mode to predict the streamflow of the Cassidy watershed —being of similar basin area. Figure 5-12 presents the results of this application. The model was very good in predicting the flow hydrograph in 2002 (a relatively dry year), but did not replicate the spring snowmelt of 2003. All 2004 hydrograph patterns were picked by the model; however, peak responses were not as good. Goodness-of-fit statistics were as follows: RB, -41%; RMSE, 0.06 m^3/s; MAE, 0.02 m^3/s; R^2, 0.56; d_2, 0.8; and r^2, 0.66. The initial results from this example are very promising. Fair prediction of three years of streamflow was achieved for an ungauged watershed using a model not trained, even with a single data point, for that specific watershed. Given that the 1A watershed has a very specific-wetland dominated soil (Prepas et al. 2006; Nour et al. 2006a) and that the originally developed model for 1A was not trained with a similar snowmelt event, it is believed that the approach would produce significantly better results if the similarity between the watershed used in model formulation and the ungauged watershed were higher in terms of basin area, soil, and vegetation type (i.e. more hydrologicaly homogenous watersheds). Yet, even with the two dissimilar watersheds, the approach is very appealing.

5.6. Scaling-up and Regionalization of Models

There is an interest to model ungauged watersheds, where no streamflow monitoring is taking place. However, the variability in climate, basin characteristics in terms of topography, vegetation, land use, and surficial geology is huge. Such variability makes it difficult in some cases to the extent that Linden and Woo (2003) had problems even when transferring a model calibrated for a basin to its sub-basins. All what we know is that the probability of having success in transferring models from one watershed to another increases when they are more hydrologicaly similar. Crude measures of hydrologic similarity that use annual water budgets can be found in the literature (e.g., Gan and Burges 2006). In many instances, even quantifying simple water budget components in an ungauged watershed is impossible due to lack of pertinent data. In this study, we proposed a measure of hydrologic similarity that relies on remote sensing (RS) information available for the public via the National Oceanic and Aeronautics Administration (NOAA). Bi-weekly composites of 250 m x 250 m pixel resolution satellite images acquired by the Moderate-resolution Imaging Spectroradiometer (MODIS) were downloaded and averaged over the area of each of the studied watersheds. Reflection information in the near infrared (λ_{NIR}) and the mid infrared (λ_{MIR}) frequency ranges were used to calculate a vegetation index termed shortwave mid infrared (SWMIR) vegetation index as explained by Eq. 15. This index was recently found to be highly correlated to leaf water content (Chen et al. 2005). In this study, we proposed a hydrologic similarity index (*SWMIR_SI*$_{i,j}$) that makes use of RS SWMIR vegetation index (Eq. 16).

$$[15] \quad SWMIR = \frac{\lambda_{NIR} - \lambda_{MIR}}{\lambda_{NIR} + \lambda_{MIR}}$$

$$[16] \quad SWMIR\_SI_{i,j} = \sqrt{\sum_{k=1}^{N} \left(\frac{SWMIR_{i,k} - SWMIR_{j,k}}{\sigma_{SWMIR}} \right)^2} \, ,$$

where λ_{NIR} and λ_{MIR} are MODIS reflection at the near infrared and the mid infrared frequency bands, respectively; *SWMIR* is the short-wave mid infrared water index at any instant of time; *SWMIR_SI*$_{i,j}$ is the proposed similarity index between basin *i* and basin *j*; *k* is a time index of two-week interval; and *N* is the number of two-week intervals in the total study duration.

A fifth watershed, the Mosquito watershed (Figure 5-1) of 3.1 km^2 basin area was used in order to test the proposed hydrologic similarity indicator. Low values of a good indicator should reveal more hydrologicaly similar basins, and thus, model performance is expected to be higher for lower values of a good hydrologic similarity indicator. All previously calibrated models were applied to the Mosquito watershed and models' performance in terms of "goodness-of fit" statistics was monitored. *SWMIR_SI*$_{i,j}$ was then calculated for each pair of watersheds. The obtained "goodness-of fit" statistics were finally regressed to the corresponding $SWMIR\_SI_{i,j}$ to assess the success of the proposed indicator in describing hydrologic similarity between the previously trained watersheds and the Mosquito watershed.

In order to account for the large variation in basin areas, two modelling approaches were implemented: first, the previously calibrated models were run on the Mosquito watershed and the predicted streamflow values were multiplied by $\text{Area}_{\text{Mosquito}}/\text{Area}_{\text{Original basin}}$ (Case 1); second, the original models were retrained by changing all the rainfall inputs to the corresponding "$\text{Area}_{\text{original basin}}.\text{Rainfall}_{\text{original basin}}$" then the ANN input layer scaling function was rescaled to accommodate the new range of "$\text{Area}_{\text{Mosquito}}.\text{Rainfall}_{\text{Mosquito}}$" before applying the models to the Mosquito watershed (Case 2). Table 5-4 summarizes the results of this application. Since R^2 was negative for some cases, it is difficult to interpret, and thus, regressing $SWMIR\_SI_{i,j}$ to a "goodness-of-fit" measure was limited to d_2 and r^2. Case 1 models were found to always behave better than case 2 models in terms of prediction accuracy reflecting the superiority of the first modelling approach over the second one (Table 5-4). Figure 5-13 summarizes the results of the conducted linear regression. Significant correlation ($r^2 > 0.71$) was established suggesting the usefulness of the proposed indicator. However, more data is required to strengthen our results.

The model with the lowest $SWMIR\_SI_{i,j}$ value—the Willow model applied on the Mosquito watershed—performed fairly well acknowledging that the calibrated watershed is 5 times bigger than the ungauged watershed (Figure 5-14).

5.7. Conclusions and Recommendations

Most of the reviewed streamflow neural network models were either recurrent network based or feed-forward multi-layer perceptron (FF-MLP) requiring the past flow values for lead-time prediction. These models cannot be used in modelling ungauged watersheds when such information is missing. The current study proposed a FF-MLP algorithm using low-cost, readily available meteorological data and careful time series manipulation prior to model building. The proposed algorithm used inverse distance weighted interpolation for better rainfall representation. The temperature index snowmelt approach was used to account for the snowmelt. Cross correlation analysis was used to identify the time-lagged inputs, and spectral analysis was used to feed the model with information representing *Q/R* hystereses loops and the flow's seasonal cyclic behaviour.

The algorithm was applied to four watersheds in the Canadian Boreal Plain. All models managed to simulate streamflow fairly well at all data ranges. Six measures of "goodness-of-fit" were used to assure model accuracy. In all cases, the best network architecture was a FF-MLP ANN with a single hidden layer. The hidden layer neurons were operating with three different activation functions. Likely, this division was analogous to the three main driving forces of streamflow (the base flow, snowmelt, and rainfall events).

To demonstrate the approach's applicability to modelling ungauged watersheds, the calibrated models were applied to a smaller watershed, the Mosquito watershed. In addition, the model initially developed for the 1A watershed (5.1 km^2) was used in a predictive mode to simulate three years of streamflow for the Cassidy watershed (5.9

km^2). The initial results from these applications are very promising. The prediction accuracy was fair in all years except in predicting the early snowmelt in 2003. A new hydrologic similarity index (*SWMIR_SI*$_{i,j}$) that makes use of public domain remote sensing information was proposed and was found to be significantly correlated to model performance. However more data is needed to strengthen our results.

5.8. References

Agarwal, A. and Singh, R.D. 2004. Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. Water Resour. Manag. **18**: 285-300.

Anctil, F., Michel, C., Perrin, C. and Andreassian, V. 2004. A soil moisture index as an auxiliary ANN input for streamflow forecasting. J. Hydrol. **286**: 155-167.

Anctill, F.and Rat, A. 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. **10**(1): 85-88.

ASCE Task Committee. 2000a. Artificial neural network in hydrology – I: preliminary concepts. J. Hydrol. Eng. **5**(2): 115-123.

ASCE Task Committee. 2000b Artificial neural network in hydrology – II: hydrologic applications. J. Hydrol. Eng. **5**(2): 124-137.

Baratti, R., Cannas. B., Fanni, A., Pintus, M., Sechi, G.M. and Toreno, N. 2003. River flow forecast for reservoir management through neural networks. Neurocomputing. **55**: 421-437.

Cahng, L.C., Chang, F.J. and Chiang, Y.M. 2004. A two-step-ahead recurrent neural network for stream-flow forecasting. Hydrol. Process. **18**(1): 81-92.

Castellano-Mendez, M., Gonzalez-Manteiga, W., Febrero-Bande, M., Prada-Sanchez, J. and Lozano-Calderon, R. 2004. Modelling of monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. J. Hydrol. **296**: 38-58.

Chang, F.J., Cahng, L.C. and Huang, H.L. 2002. Real-time recurrent learning neural network for stream-flow forecasting. Hydrol. Process. **16**(13): 2577-2588.

Chen, D., Huang, J. and Jackson, T.J. 2005. Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. Remote Sens. Environ. **98**: 225-236.

Dingman, S.L. 2002. Physical Hydrology, second ed. Prentice Hall, Upper Saddle River, New Jersey.

Fox, D.G. 1981. Judging air quality model performance: a summary of the AMS workshop on dispersion model performance. Bull. Am. Meteorol. Soc. **62**: 599-609.

Gamal El-Din, A and Smith, D.W. 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. **36**(5): 1115-1126.

Gan, T.Y. and Burges, S.J. 2006. Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. J. Hydrol. **320**: 117-131.

Garson, G.D. 1991. Interpreting neural-network connection weights. AI Expert 4: 47-51.

Haykin, S. 1994. Neural networks: a comprehensive foundation, second ed. Macmillian CollegePublishing, Inc., New York..

Kisi, O. 2004. River flow modelling using artificial neural networks. J. Hydrol. Eng. **9**(1): 60-63.

Kumar, D.N., Raju, K.S. and Sathish, T. 2004. River flow forecasting using recurrent neural networks. Water. Resour. Manag. **18**: 143-161.

Legates, D.R. and McCabe Jr., G.J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water. Resour. Res. **35**(1): 233-241.

Linden, S.V.D. and Woo, M. 2003. Transferability of hydrological model parameters between basins in data-sparse areas, sub-Arctic Canada. J. Hydrol. **270**: 182-194.

Maier, H.R.and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw.**15**: 101-124.

Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, part 1: a discussion of principles. J. Hydrol.**10**: 282-290.

Nour, M.H., Smith, D.W., Gamal El-Din, M. and Prepas, E.E. 2006a. The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. Ecol. Model. **191**: 19-32.

Nour, M.H., Smith, D.W., Gamal El-Din, M. and Prepas, E.E. 2006b. Artificial neural networks and time series modelling of TP concentration in boreal streams: a comparative approach. J. Environ. Eng. Sci. **5**(S1): S39-S52.

Nour, M.H., Smith, D.W., Gamal El-Din, M. and Prepas, E.E. 2005. Towards a generic artificial neural networks model for dynamic predictions of stream flow in ungauged watersheds. Proceedings of the Eighth International Conference on the Application of Artificial Intelligence to Civil, Structural and Environmental Engineering. B.H.V. Topping, (Editor), Civil-Comp Press, Stirling, United Kingdom, paper 49.

Prepas, E.E., Burke, J.M., Whitson, I.R., Putz, G. and Smith, D.W. 2006. Associations between watershed characteristics, runoff, and stream water quality: hypothesis development for watershed disturbance experiments and modelling in the Forest Watershed and Riparian Disturbance (FORWARD) project. J. Environ. Eng. Sci. **5**(S1): S27-S37.

Riad, S., Mania, J., Bouchaou, L. and Najjar, Y. 2004. Predicting catchement flow in a semi-arid region via an artificial neural network technique. Hydrol. Process. **18**(13): 2387-2393.

Riad, S.and Mania, J. 2004. Rainfall-Runoff model using an artificial neural network approach. Math. Comput. Model.**40**: 839-846.

Sudheer, K.P., Nayak, P.C. and Ramasastri, K.S. 2003. Improving peak flow estimates in artificial neural network river flow models. Hydrol. Process. **17**(3): 677-686.

Tawfik, M. 2003. Linearity versus non-linearity in forecasting Nile River flows. Adv. Eng. Softw. **34**: 515-524.

Tokar, A.S. and Markus, M. 2000. Precipitation-Runoff modelling using artificial neural networks and conceptual models. J. Hydrol. Eng. **5**(2): 156-161.

Ward Systems Group. 1996. NeuroShell 2 version 3 reference manual, Inc., Frederick, Maryland, USA.

Willmott, C.J. 1981. On the validation of models. Phys. Geogr. **2**: 184-194.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J. and Rowe, C.M. 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. **90**: 8995-9005.

Table 5-1. Summary table for all models' inputs

Model	Inputs
1A Model	R_t, R_{t-1}, R_{t-2}, R_{t-3}, $sin(2\pi vt)$, $cos(2\pi vt)$, T_{max}, T_{min}, dd_t, S_t
Cassidy Model	R_t, R_{t-1}, $sin(2\pi vt)$, $cos(2\pi vt)$, T_{max}, T_{min}, dd_t, S_t
Two Creek Model	R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, $sin(2\pi vt)$, $cos(2\pi vt)$, T_{max}, T_{min}, dd_t, S_t
Willow Model	R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, $sin(2\pi vt)$, $cos(2\pi vt)$, T_{max}, T_{min}, dd_t, S_t

where: R_t through R_{t-6} are the estimated rainfall at the centroid of the watershed in mm/d at lags 0 through 6; T_{max} and T_{min} represents maximum and minimum air temperatures in °C, respectively; dd_t, is the cumulative degree days; and S_t, is the cumulative snowfall in cm.

Table 5-2. Summary table showing optimum ANN models' architecture and ANN internal parameters

	1A Model	Cassidy Model	Two Creek Model	Willow Model
Data division (TS:SS:CVS)	3:1:1			
Scaling function	Linear, <-1,1>	Linear, <-1,1>	Linear, <-1,1>	Linear, <-1,1>
Optimum network (I-[H-H-H]-O)	10L-[7G-7GC-5LO]-T	8L-[5G-5GC-5LO]-LO	13L-[5G-5GC-5T]-LO	13-L-[4G-4GC-4LO]-LO
Training algorithm	BP-BU			
Learning rate	Insensitive			
Momentum coefficient	Insensitive			
Initial weights	Random [-0.3,0.3]			
Epoch size	TS (485)	TS (351)	TS (480)	TS (383)
Stopping criterion	Best test set (in terms of MSE)			

where: I and O denote input and output layers, respectively.; [H-H-H], represents a single hidden layer with different activation function; L, is the linear scaling function; G, GC, LO, and T are the Gaussian, Gaussian complement, logistic, and the hyperbolic tan activation functions, respectively; TS, SS, and CVS are the training, the testing, and the cross-validation data sets, respectively; and < > denotes an open interval.

Table 5-3. Statistical measures of models' performance

	1A Model			Cassidy Model			Two Creek Model			Willow Model		
	TS	SS	CVS	TS	SS	CVS	TS	SS	CVS	TS	SS	CVS
RB(%)	-7	-9	-18	1	9	14	1	-7	-3	5	7	11
RMSE	0.05	0.08	0.06	0.01	0.03	0.05	0.52	1.19	1.59	0.05	0.08	0.08
MAE	0.03	0.04	0.03	0.01	0.02	0.02	0.29	0.69	0.75	0.02	0.04	0.04
R^2	0.91	0.79	0.75	0.98	0.90	0.78	0.97	0.77	0.71	0.92	0.71	0.79
d_2	0.98	0.94	0.92	0.99	0.97	0.95	0.99	0.93	0.92	0.98	0.92	0.95
r^2	0.91	0.8	0.76	0.98	0.90	0.86	0.97	0.77	0.72	0.96	0.72	0.82

TS, training data set; SS, testing data set; CVS, cross-validation data set; RMSE and MAE are in m^3/s

Table 5-4. Statistical measures of models' performance applied to the Mosquito watershed

	1A model applied to Mosquito		Two Creek model applied to Mosquito		Cassidy model applied to Mosquito		Willow model applied to Mosquito	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
RB(%)	99	-39	179	5	22	-36	24	-27
RMSE	0.04	0.03	0.04	0.03	0.02	0.02	0.02	0.02
MAE	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01
R^2	-2.06	-0.08	-1.97	-0.32	0.33	0.46	0.57	0.51
d_2	0.58	0.54	0.64	0.45	0.82	0.76	0.88	0.79
r^2	0.27	0.13	0.4	0.05	0.48	0.49	0.63	0.55

Class 1 models are original models applied to the Mosquito watershed and the output was then scaled to reflect Mosquito's basin area, Class 2 models are original models retrained using "Area x Rain" as inputs instead of Rain values then applied to the Mosquito watershed, and RMSE and MAE are in m^3/s

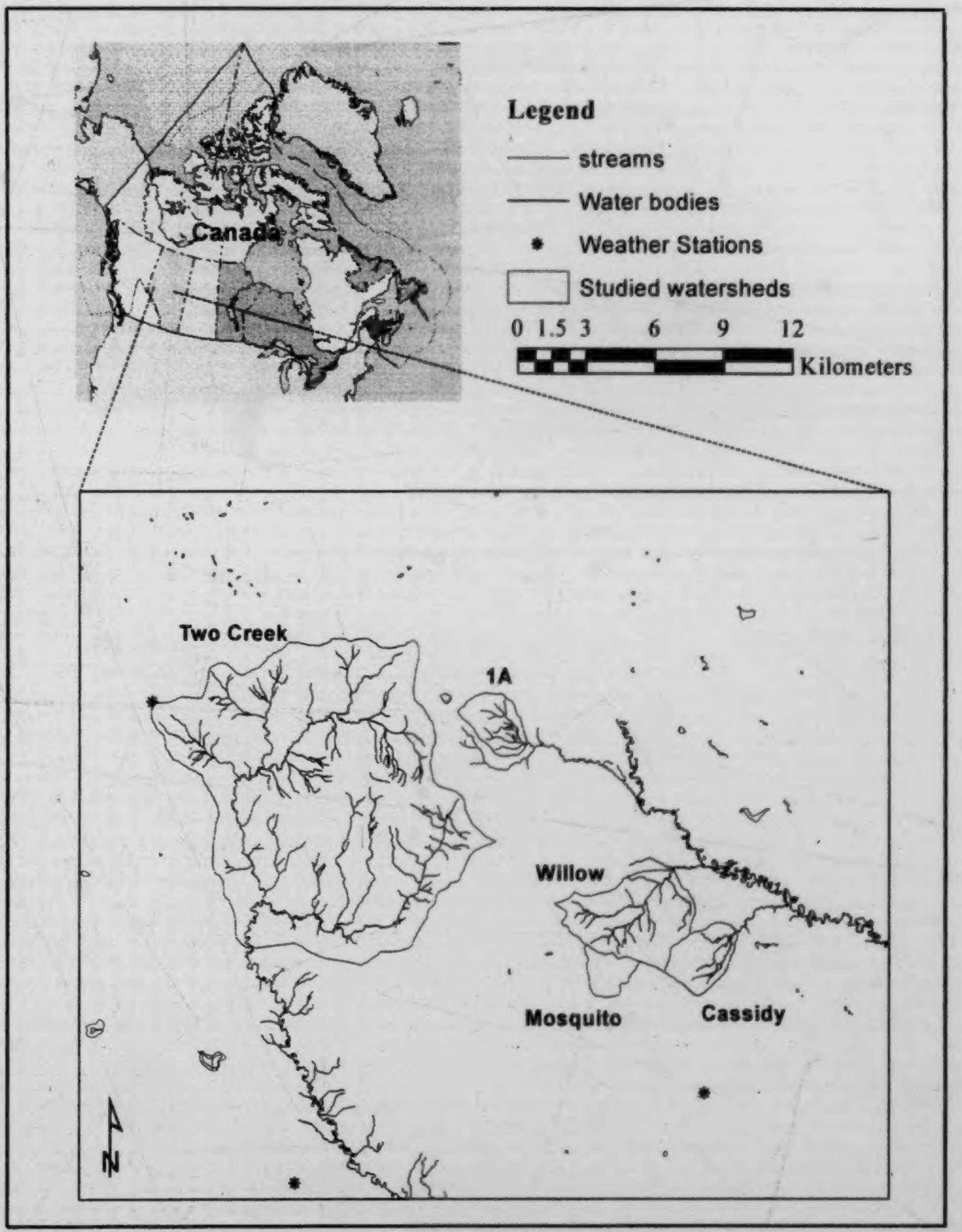

Canada
Legend
streams
Water bodies
Weather Stations
Studied watersheds
0 1.5 3 6 9 12
Kilometers
Two Creek
1A
Willow
Mosquito
Cassidy
N

Figure 5-1. Study area showing studied watersheds and utilized weather stations (some weather stations are not shown at this scale)

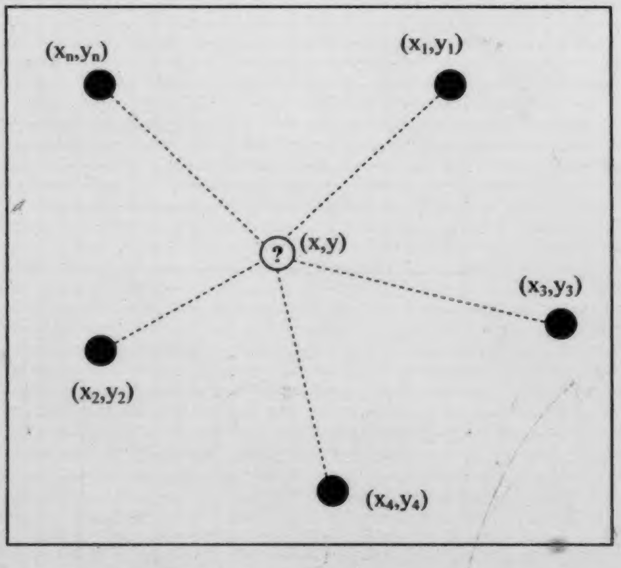

(x_n,y_n)
(x_1,y_1)
?
(x,y)
(x_3,y_3)
(x_2,y_2)
(x_4,y_4)

Figure 5-2. Schematic of the inverse distance weighted (IDW) interpolation

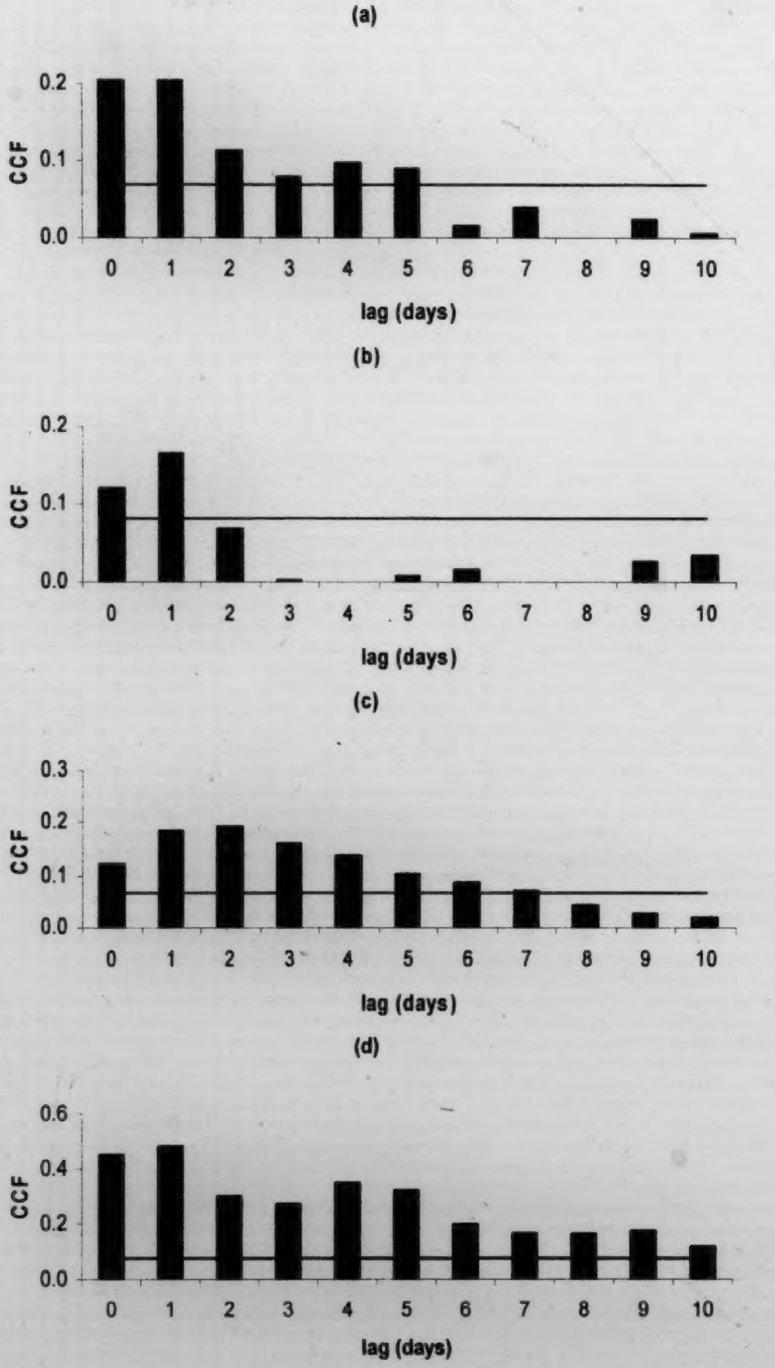


Figure 5-3. CCF (*Q*, *R*) for (a) 1A, (b) Cassidy, (c) Two Creek, and (d) Willow watersheds. Solid line shows the 95% confidence boundary

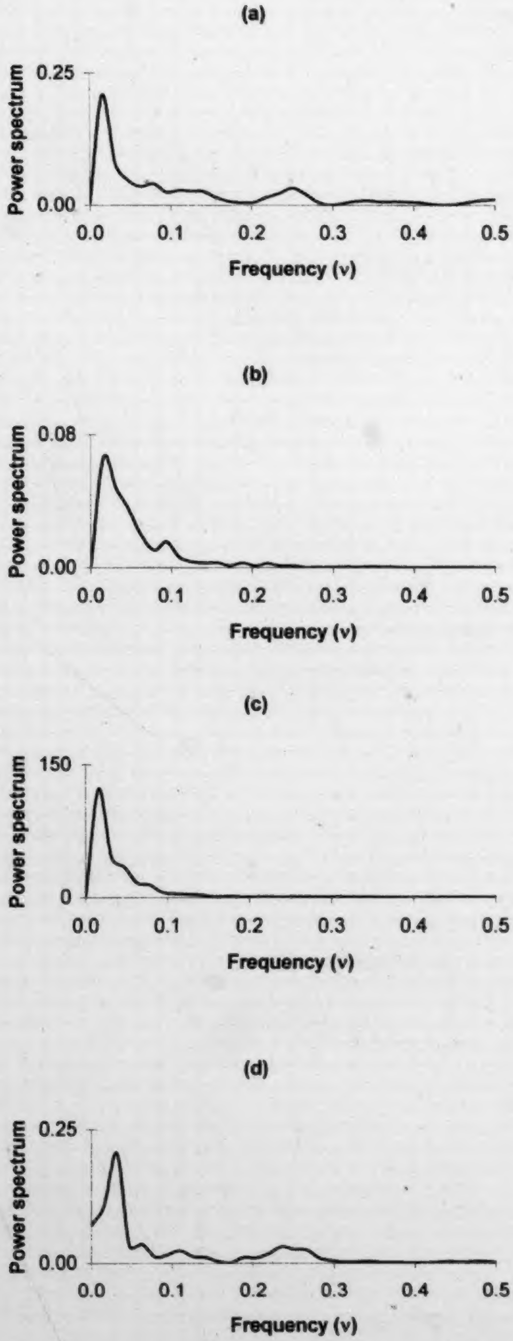


Figure 5-4. Power spectrum of flow time series in $[m^3/s]^2$ for (a) 1A, (b) Cassidy, (c) Two Creek, and (d) Willow watersheds

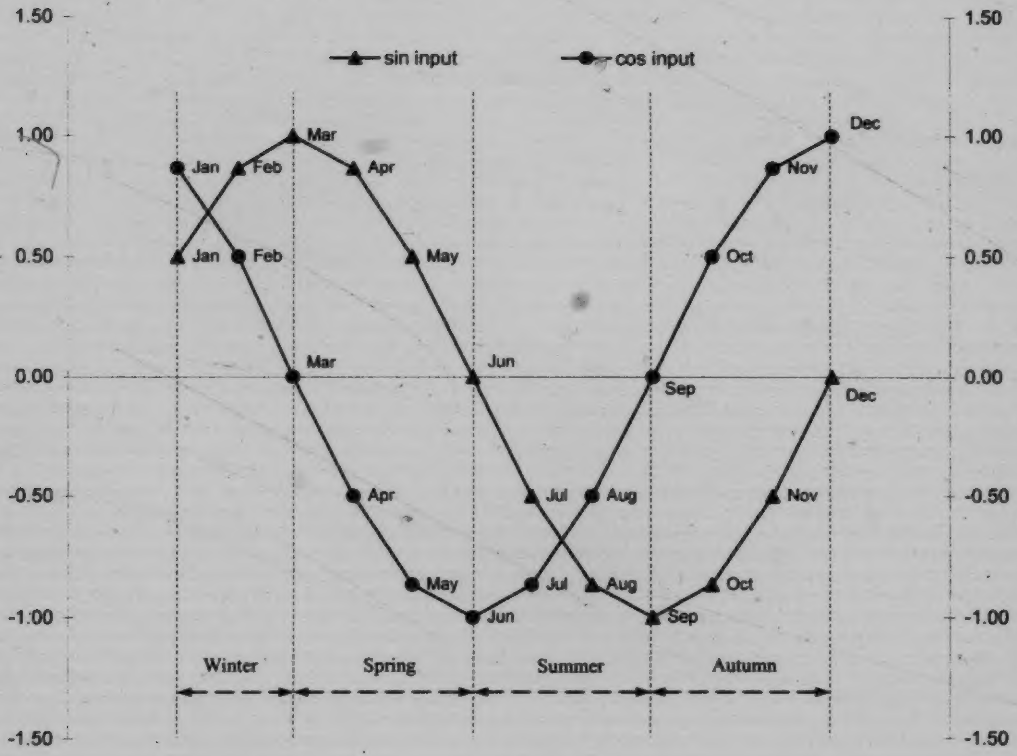


Figure 5-5. A description of the concept of feeding ANN *Q* models with seasonal variation and *Q*/*R* hystereses loops

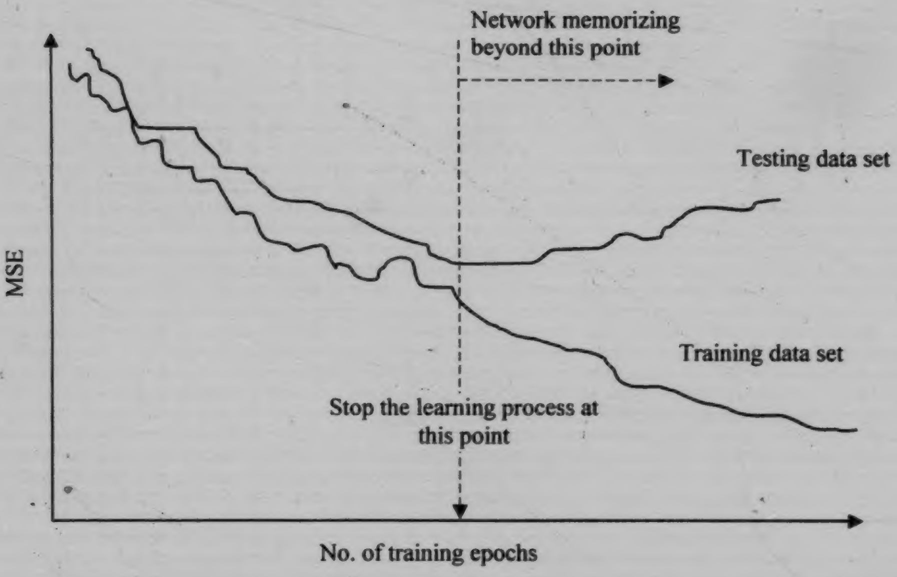

Network memorizing beyond this point
Testing data set
MSE
Training data set
Stop the learning process at this point
No. of training epochs

Figure 5-6. Schematic showing the usage of the testing data set in optimizing model training

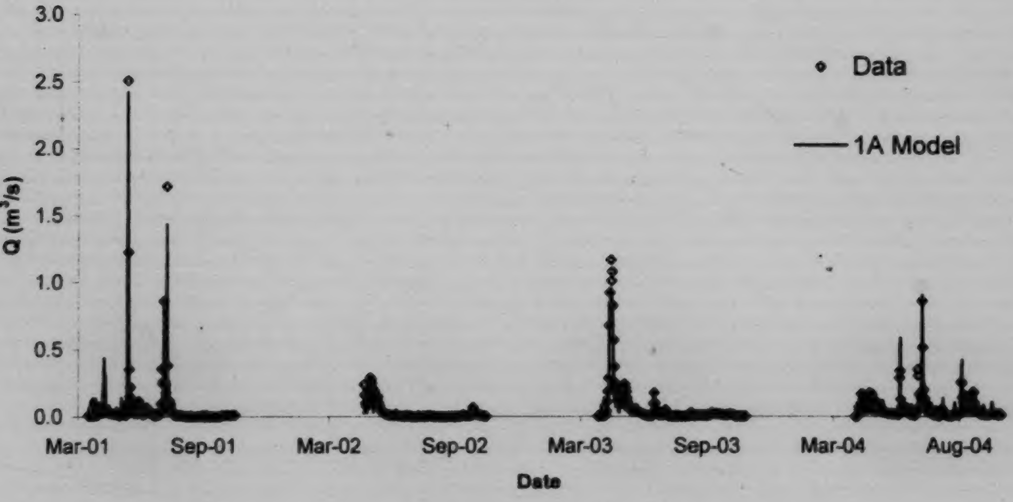


Figure 5-7. Comparison of measured and modelled flow hydrographs of 1A watershed

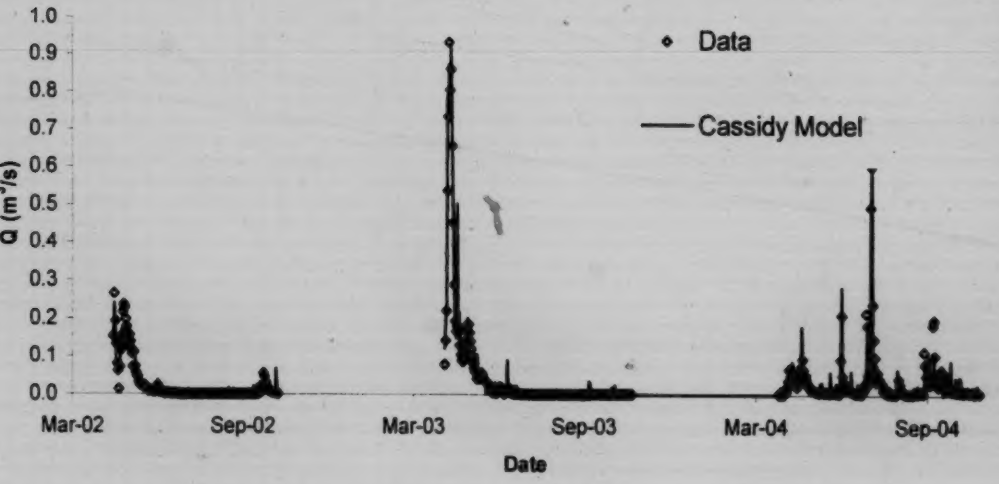


Figure 5-8. Comparison of measured and modelled flow hydrographs of Cassidy watershed

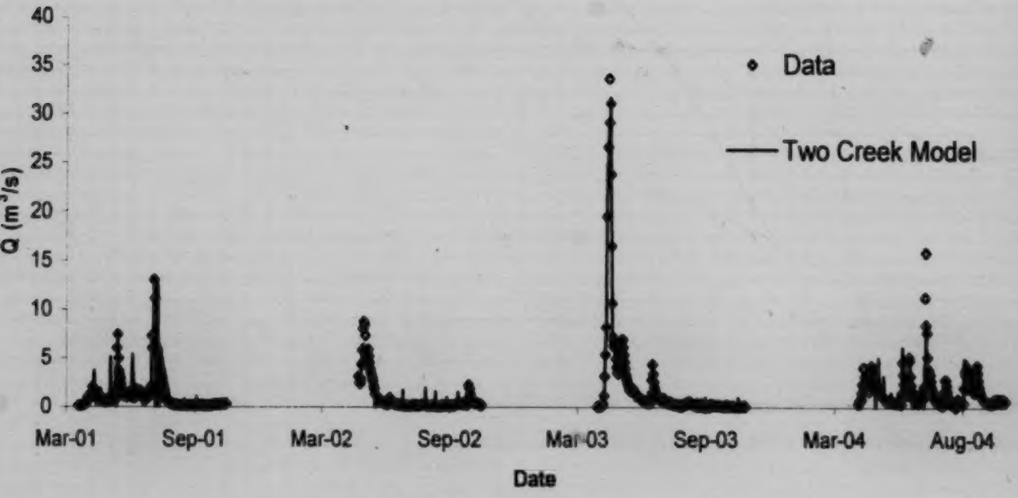


Figure 5-9. Comparison of measured and modelled flow hydrographs of Two Creek watershed

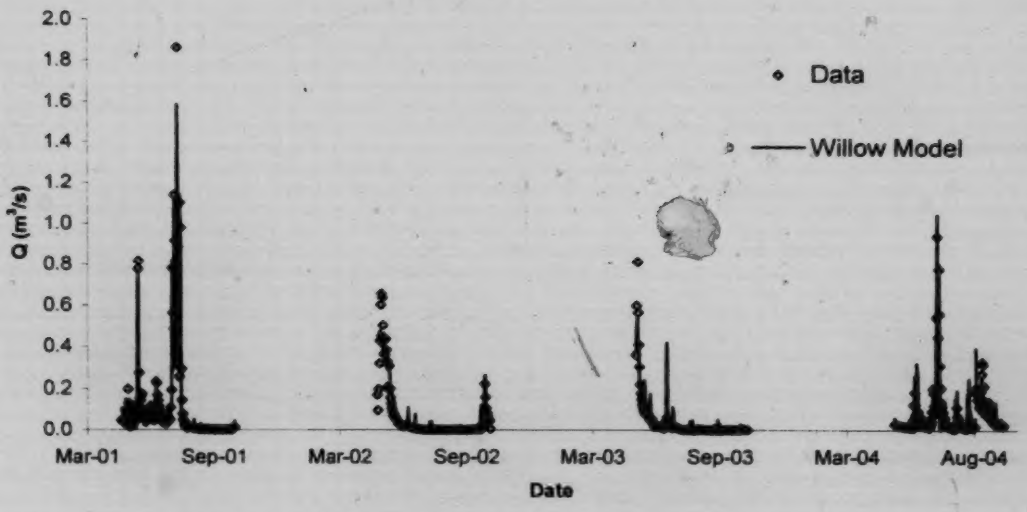


Figure 5-10. Comparison of measured and modelled flow hydrographs of Willow watershed

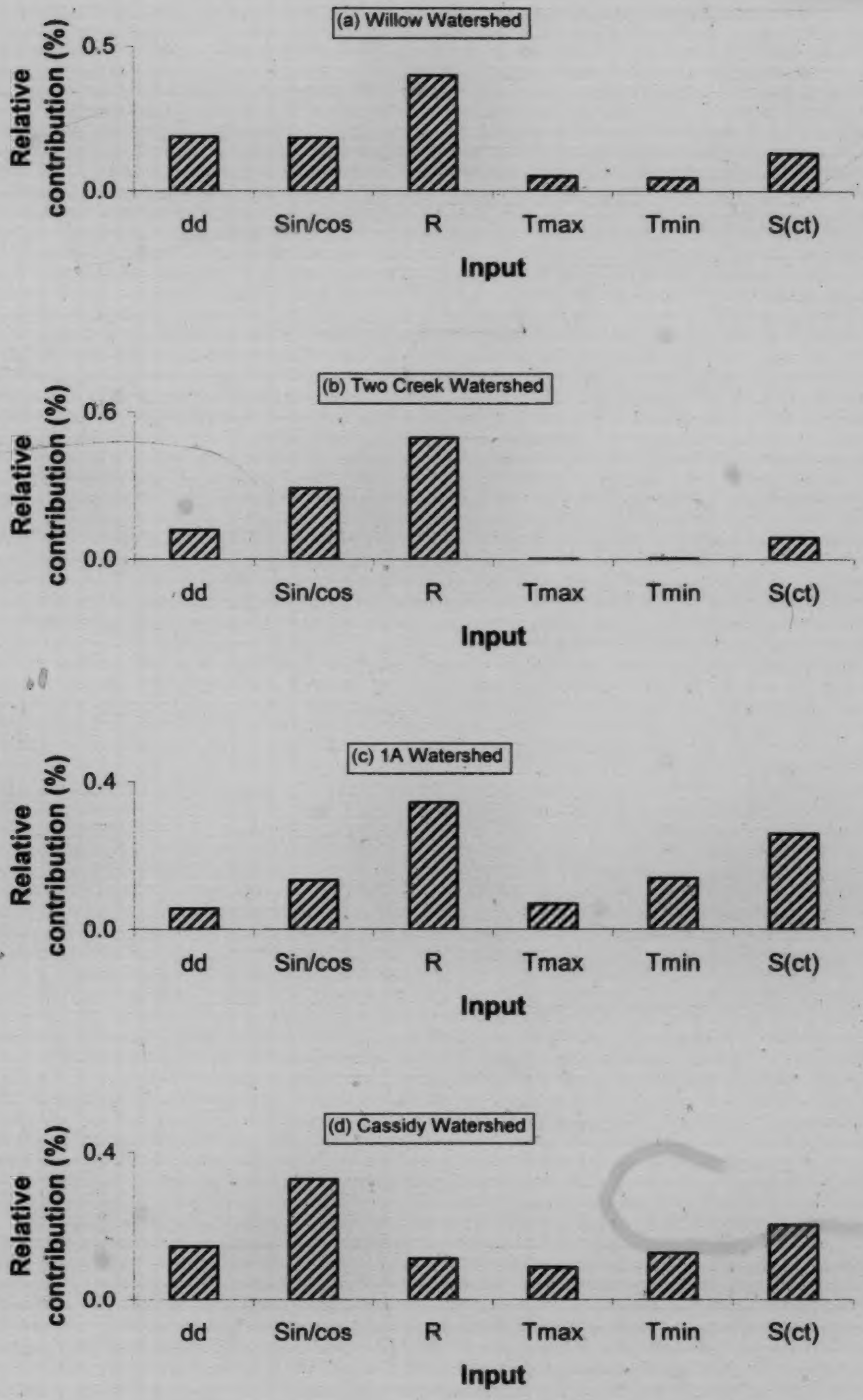


Figure 5-11. Importance of model inputs for streamflow predictions

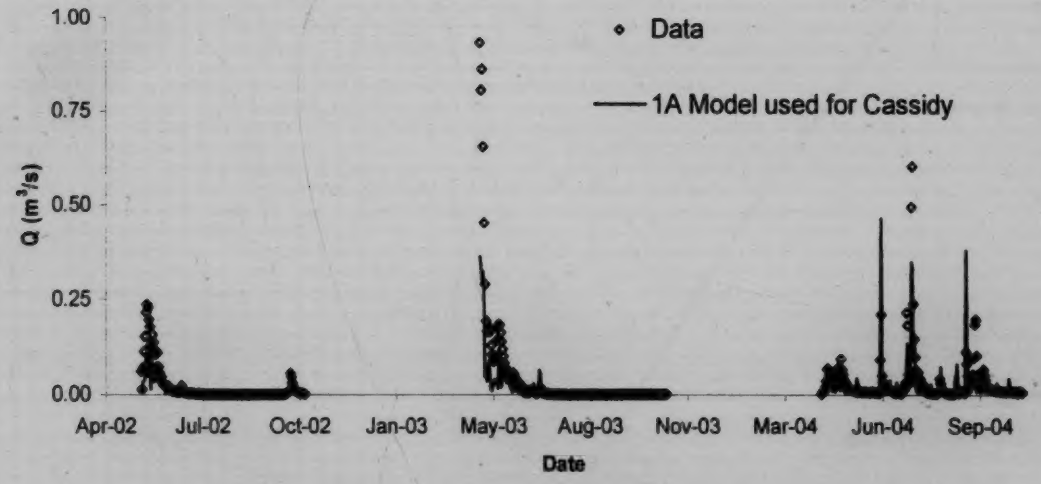


Figure 5-12. Comparison of measured and modelled flow hydrographs of Cassidy watershed when 1A ANN model was used for Cassidy flow prediction

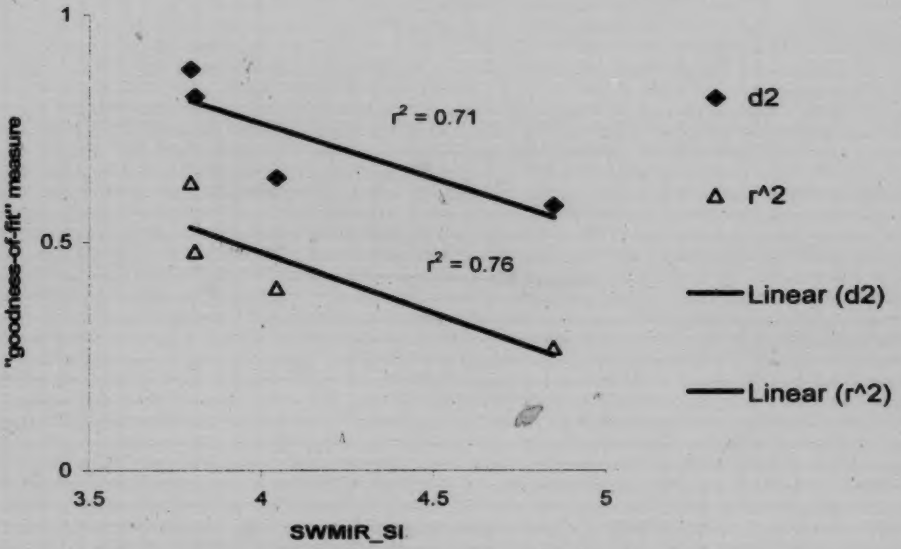


Figure 5-13. A simple linear regression between models' "goodness-of-fit" measures and SWMIR_SI similarity index

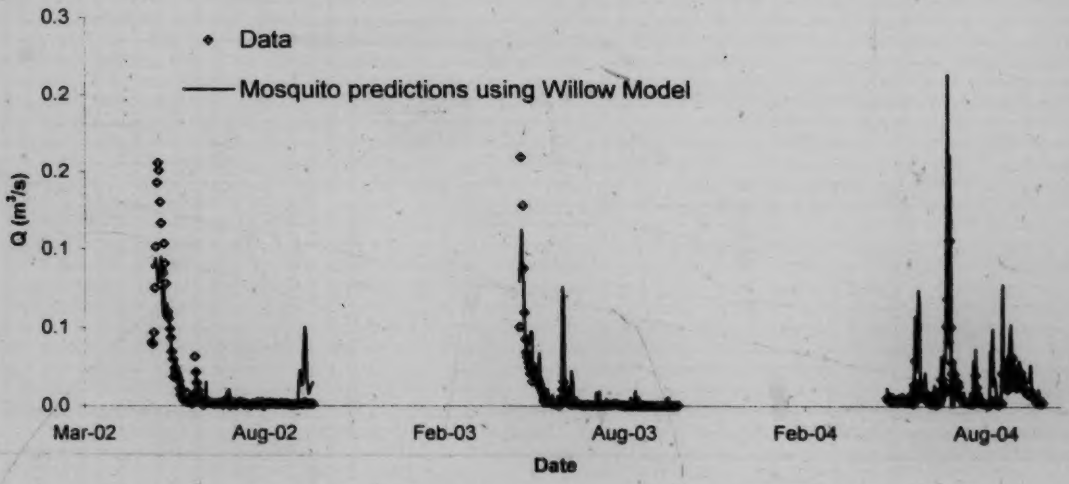


Figure 5-14. Comparison of measured and modelled flow hydrographs of Mosquito watershed when Willow ANN model (class 1) was used for Mosquito flow predictions

CHAPTER 6. DEVELOPING A TRAIN OF ARTIFICIAL NEURAL NETWORKS MODELS FOR MODELLING STREAMFLOW, SUSPENDED SOLIDS, AND PHOSPHORUS

6.1. Introduction

During storm events and snow melts—when the soil is susceptible to erosion—not only does total suspended solids (*TSS*) increase considerably but it can also act as a vector for the transport of many contaminants like; phosphorus, nitrogen, organic matter, and heavy metals (Gong et al. 1996; Munn and Prepas 1986). The resulting increase in nutrient loading to receiving water bodies, particularly phosphorus, can potentially lead to an increase in algal biomass, dissolved oxygen depletion, cyanobacteria toxin production, and ultimately accelerated eutrophication (Carpenter et al. 1998). Thus a reasonable prediction of *TSS*, *TP*, and their associated *Q* is critical to preserving aquatic ecosystems of interest.

Although conceptual models—for *Q*, *TSS*, and *TP*—are considered to be the best in terms of providing an understanding of the hydrological, geological, biogeochemical, and physical processes controlling water and contaminant transport, they are undermined in practice due to the incomplete understanding of the transport mechanisms at the watershed scale, and for the scarcity of data required for model calibration.

A version of this chapter has been published. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas E.E. 2006. Neural networks modelling of streamflow, phosphorus, and suspended solids: application to the Canadian Boreal forest. *Water Sci. Technol.* **53**(10): 91-99.

Alternatively, artificial neural network (ANN) models have gained popularity for efficient modelling of non-linear systems—like the one under investigation. ANNs have been successfully applied to modelling streamflow, salinity, pH, and other water quality variables (Maier and Dandy 2000; Brion and Lingireddly 2003; Lek et al. 1996; Wilson and Recknagel 2001; Moatar et al. 1999; Zhang and Stanley 1997; Zhang et al. 2004) providing premise for utilization in the current study.

The objectives of this study were to develop robust ANN models for modelling daily streamflow, total suspended solids, and total phosphorus; to apply the developed models to a 130-km^2 watershed in the Canadian Boreal forest; and to provide a framework for model implementation into water resources management plans.

6.2. Materials and Methods

The daily Q, TSS, and TP concentrations were collected as part of the forest watershed and riparian disturbance (FORWARD) project for the Two Creek watershed (130 km^2; latitude, 54.4° N; longitude, -116.4° W) from 2001 to 2004 (Detailed description of the project and data collection protocols can be found on *J. Environ. Eng. Sci.* special issue Volume 2, 2003). The daily rainfall (R) was obtained from 15 fire towers and weather stations in close proximity to the study watershed. The snowfall (S), mean, maximum, and minimum air temperatures (T_{mean}, T_{max}, and T_{min}, respectively) were recorded at the Environment Canada Whitecourt weather station (latitude, 54.15°; longitude, -115.78°).

The rainfall values recorded at each weather station were significantly different (up to 50% variation in some cases). Thus, in order to obtain a more representative rainfall time series for the study watershed, the inverse distance weighted (IDW) interpolation scheme was utilized (Detailed description of the approach and its implementation is presented in Weber and Englund (1992)). In the ensuing sections, R_t will be used to denote the IDW interpolated rainfall at time t. The mean air temperature and snowfall obtained from the Environment Canada weather station were further processed to calculate the cumulative degree-days (dd_t) and the cumulative snowfall (Sc_t) at time t (Equations are presented in Nour *et al.* 2006a). The Sc_t was used to reflect the available snow depth, representing the amount of snow available for melting, and the dd_t was used to provide an integrated measure of the heat energy available for snow melting.

In the current study, a streamflow ANN model was developed to predict Q from meteorological information. Modelled Q values were then coupled with weather information to serve as inputs for a *TSS* ANN model. Finally, modelled Q and *TSS* were used to augment weather data in predicting *TP* concentration. The utilized approach provided robust modelling of all three parameters utilizing data currently available to the public from Environment Canada at no cost.

6.3. ANN Model Development

Among many ANN topologies, the feed-forward multilayer perceptron (FF-MLP) ANN is by far the most widely used in water quantity and quality studies (Maier and Dandy 2000). FF-MLP is a neural network constituted by an input layer, one or more hidden layers, and an output layer. The nodes of any two consecutive layers are mutually connected but information only flows in one direction with no feedbacks (More detailed descriptions of ANNs are given in many journal papers and textbooks e.g. Haykin (1994)). FF-MLP networks are typically trained by supervised learning using the back-propagation (BP) algorithm (Haykin 1994). Unlike FF-MLP ANN, Kohonen neural network (KNNs), sometimes referred to as Kohonen self-organizing maps (SOMs), are based on the unsupervised learning methodology, in which the relevant multivariate algorithms seek clusters in the data. Unsupervised learning allows the investigator to group objects together based on their perceived similarity (Kohonen 1982). During the training process, the data set, which consists of a large number of patterns, is presented to the SOM. The Euclidean distances between a pattern and all output neuron patterns are calculated. The weights of the output neuron, which is the nearest to the pattern, and also of its neighbouring neurons, are modified so that it moves slightly closer to the input pattern. After an iterative training process, neighbouring output neurons will have more similar patterns than distant ones. In this way, a Kohonen layer consisting of homogeneous groups of data can be obtained (Figure 6-1).

This study utilized the capabilities of KNN in pattern recognition to divide the data into training (for model calibration) and testing (for model validation) data sets. FF-MLP was used for ANN model formulation. The main building blocks in model construction are described below.

6.3.1. Determination of Model Inputs

Adequate understanding of the processes deriving the modelled parameter is critical to identifying model causal inputs (factors that are known to affect the modelled parameter). However, when modelling time correlated variables, in addition to causal inputs, additional inputs should be included to reflect the time dependence and the seasonal cyclic nature of the modelled variables.

This study modelled *Q*, *TSS*, and *TP* utilizing information available to the public via Environment Canada without the need for supplementary case-specific information. The rationale behind the choice of cause/effect type inputs was to identify a surrogate for each component of the water cycle with emphasis on components representing storm events and snow melts. However, to be able to apply such modelling tool in water resources management of the region, all inputs should be easily acquired at a reasonable cost (preferably to be available for the public at no cost).

Cross correlation analysis was used to evaluate the strength of the relation between the modelled variable(s) and the potential time-lagged input variables. The time-lagged inputs that proved to be correlated to the modelled parameter(s) were identified as additional model inputs. Spectral analysis was utilized to quantify data periodicity in terms of the dominant frequency (ν) (see Shumway and Stoffer (2000) for details). This information was then used—in terms of two additional model inputs ($\sin(2\pi\nu t)$ and $\cos(2\pi\nu t)$)—to reflect the seasonal variations in the modelled parameters, *TSS*/*Q* and *TP*/*Q* hystereses loops. A summary of causal and time series inputs utilized in modelling the three parameters of interest is shown in Figure 6-2.

6.3.2. Data Division

Maier and Dandy (2000) highlighted the importance of data division in ANN modelling. Training and testing data sets should be statistically similar to achieve model robustness. KNN was used to separate available data into as many statistically homogeneous groups as possible (Figure 6-1). Implementing this approach yielded 38, 36, and 35 homogenous clusters for *Q*, *TSS*, and *TP*, respectively. Each group is then divided into training and testing data sets in the ratio of 3:1 formulating two statistically similar (homogeneous) groups; a training data set (for model calibration) and a testing data set (for model validation).

6.3.3. Determination of Network Architecture

Earlier work has proven that a typical FF-MLP ANN with a single hidden layer that utilizes one activation function in its processing elements (nodes) cannot accurately map *Q* and *TP* in higher latitudes due to the added complexity resulting from snow melts (Nour et al. 2006a,b). The complexity in modelling streamflow and its associated water quality parameters is mostly due to the dynamic change of the driving forces controlling their magnitude. At times, rainfall events are the main driving forces of these variables, but at other times, either base flow or snow melt may control their values. Thus, a modification was made by applying FF-MLP ANN with one hidden layer with processing elements that utilize more than one activation function. A combination of activation functions, scaling functions, learning and momentum rates, and training algorithms were systematically examined attempting to build the most possible parsimonious ANN models. The configurations that resulted in the best model performance (for *Q*, *TSS*, and *TP*) are illustrated in Figure 6-2.

6.4. Results and Discussion

6.4.1. ANN Q Model

The performance of each candidate model (representing different model configurations) was assessed based on statistical and graphical means. Root mean squared error (RMSE) and the coefficient of multiple determination (R^2) of the training and testing data sets were compared in each case and important data patterns were graphically examined. The model that performed the best utilized a linear scaling function that scales the inputs in the open interval of -1 and 1, one hidden layer (18 nodes; each 6 operates with one of the following activation functions: Gaussian, Gaussian complement, and tanh). The logistic activation function was used in the output layer. The best network was trained with BP algorithm with a batch update (BP-BU) technique (see Haykin (1994) for details). This training algorithm was insensitive to the learning and momentum rates. The developed model provided good flow predictions for both the training and testing data sets (RMSE values were 0.6 and 1.0 m^3/s for the training and testing data sets, respectively). Figure 6-3a is a scatter plot of modelled and measured Two Creek watershed streamflow. The model was capable of accurately mapping the flow hydrograph at all data ranges (Figure 6-3b). R^2 values were 0.95 and 0.89 for the training and testing data sets, respectively. The retained good model performance for the two data sets reflects high generalization ability.

6.4.2. ANN TSS Model

Simulated flow values were augmented with meteorological information to construct a *TSS* model. The developed model performance was even better than the flow model. The optimized network architecture was similar to the ANN flow model architecture in everything except in the hidden layer neurons. In the *TSS* model, 30 neurons were used (each 10 operates with one of the following activation functions: Gaussian, Gaussian complement, and tanh). Measured and ANN predicted *TSS* profiles were in good agreement. R^2 values exceeded 0.91 for the training and testing data sets (Figure 6-4a). RMSE of 18 and 24 mg/L were obtained for the training and testing data sets, respectively. The recorded RMSE was small as compared to the magnitude of the corresponding *TSS* values. Figure 6-4b portrays the measured and the modelled *TSS* profiles. It shows very good match of the measured and the predicted profiles with the model slightly underestimating peaks during snowmelt events. Whereas, incorporating rain and snowmelt information—as model inputs—contributed largely to mapping peak *TSS* concentration, simulated Q values likely enhanced *TSS* simulation in times when base flow was dominating.

6.4.3. ANN TP Model

Akin to the ANN *TSS* model, simulated Q and *TSS* values were used along with other meteorological data to formulate the *TP* ANN model. The optimum network configuration was identical to the ANN Q model configuration with the exception that the tanh activation function was used in place of the logistic activation function in the output layer. The simulation performance was as good as the previous two models (Figure 6-5). The (R^2 value, RMSE) pair was (0.90, 11 μg/L) and (0.89, 15 μg/L) for the training and testing data sets, respectively. The similarity in the obtained optimum network configuration for the three modelled variables highlighted the robustness of the model building approach and the possibility of generalizing the approach to other water quality parameters.

6.5. Conclusions and Final Remarks

The current study proposed an artificial neural network (ANN) modelling algorithm that relies on low-cost readily available meteorological data for modelling streamflow (*Q*), total suspended solids (*TSS*) and total phosphorus (*TP*) concentrations. The models were applied to a 130-km^2 watershed in the Canadian Boreal Plain. Our results demonstrated that through careful manipulation of time series analysis and rigorous optimization of ANN configuration, it is possible to simulate *Q*, *TSS*, and *TP* reasonably well. R^2 values exceeding 0.89 were obtained for all modelled data cases.

Models of this kind can provide very useful applications. They can provide real time prediction of *Q*, *TSS*, and *TP* in the studied watershed and can potentially be extrapolated to hydrologically similar watersheds. In both cases, questions related to climate change scenarios and their impact on streamflow and water quality can be addressed. Moreover, probabilities of exceeding a certain threshold of a contaminant can also be calculated by such models. Figure 6-6 illustrates a proposed framework for utilizing the developed models in water resources management. Of interest to water resources management is to assess the impact of rare events (like 1/20 years storm event) on water quantity and quality. This task can be demonstrated through evaluating the impact of a hypothetical 1/30 years storm event as follows: (1) the probability density function (PDF) for all ANN *Q* model inputs can be calculated (in North America, historic meteorological data at different weather stations is currently available for periods from 30 to over 100 years allowing such calculations); (2) through Monte Carlo simulations, one can randomly sample input PDFs and run the ANN *Q* model for all realizations resulting in a cumulative probability distribution (CDF) of the responsive *Q* (Figure 6-6); (3) the obtained *Q* PDF can be sampled again with other *TSS* ANN model inputs by Monte Carlo simulations and similarly, PDF and CDF can be calculated for the *TSS*; and (4) the same procedure can be repeated for the *TP* concentration. The obtained CDFs for *Q*, *TSS*, and *TP* can then be used in water resources management by estimating the probability of exceeding a certain threshold of the modelled parameters in the event that a 1/30 years event is to occur. This information can guide watershed land use activities and management plans.

6.6. References

Brion, G.M., and Lingireddy S. 2003. Artificial neural network modelling: a summary of successful applications relative to microbial water quality. *Water Sci. Technol.* **47**(3): 235-240.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., and Smith, V.H. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen.

Ecol. Appl., **8**(3): 559-568.

Gong, N., Denoeux, T., and Bertand-Krajewski, J. 1996. Neural networks for solid transport modelling in sewer systems during storm events. *Wat. Sci. Technol.* **33**(9): 85-92.

Haykin, S. 1994. Neural Networks: A Comprehensive Foundation. Macmillian College Publishing, NY.

Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps. *Biol. Cybern.* **43**(1): 59-69.

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus concentration and phosphorus load from watershed characteristics using backpropagation neural networks. Acta Oecol. **17**(1): 43-53.

Maier, H.R., and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and application. Environ. Modell. Softw. **15**: 101-124.

Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural networks. Application of control and validation data series in Middle Loire River. Ecol. Model. **120**: 141-156.

Munn, N., and Prepas, E.E. 1986. Seasonal dynamics of phosphorus partitioning and export in two streams in Alberta, Canada. Can. J. Fish. *Aquat. Sci.* **43**: 2464-2471.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006a. The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. Ecol. Model. **191**: 19-32.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006b. Artificial neural networks and time series modelling of TP concentration in boreal streams: A comparative approach. J. Environ. Eng. Sci. **5**: S39-S52.

Shumway, R.H., and Stoffer D.S. 2000. Time Series Analysis and its Applications. Springer, NY.

Weber, D., and Englund, E. 1992. Evaluation and comparison of spatial interpolators. *Math. Geol.* **24**: 381-391.

Wilson, H., and Recknagel, F. 2001. Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. Ecol. Model. **146**: 69-84.

Zhang, Q., and Stanley, S.J. 1997. Forecasting raw-water quality parameters for North Saskatchewan River by neural network modelling. Water Res. **31**: 2340-2350.

Zhang, Q.J., Cudrak, A.A., Shariff, R., and Stanley, S.J. 2004. Implementing artificial neural network models for real-time water colour forecasting in a water treatment plant. J. Environ. Eng. Sci. **3**: S15-S23.

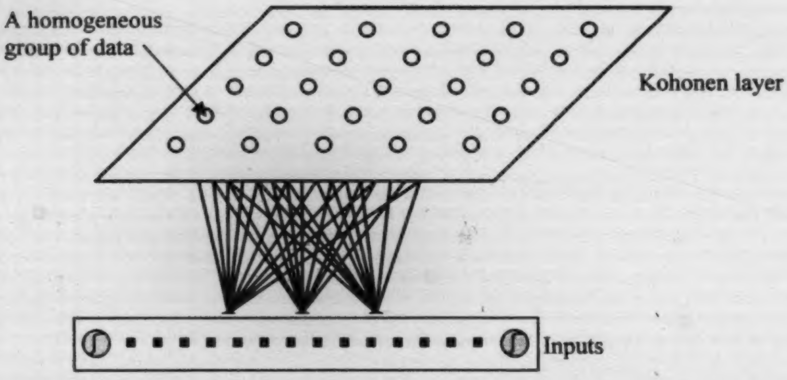
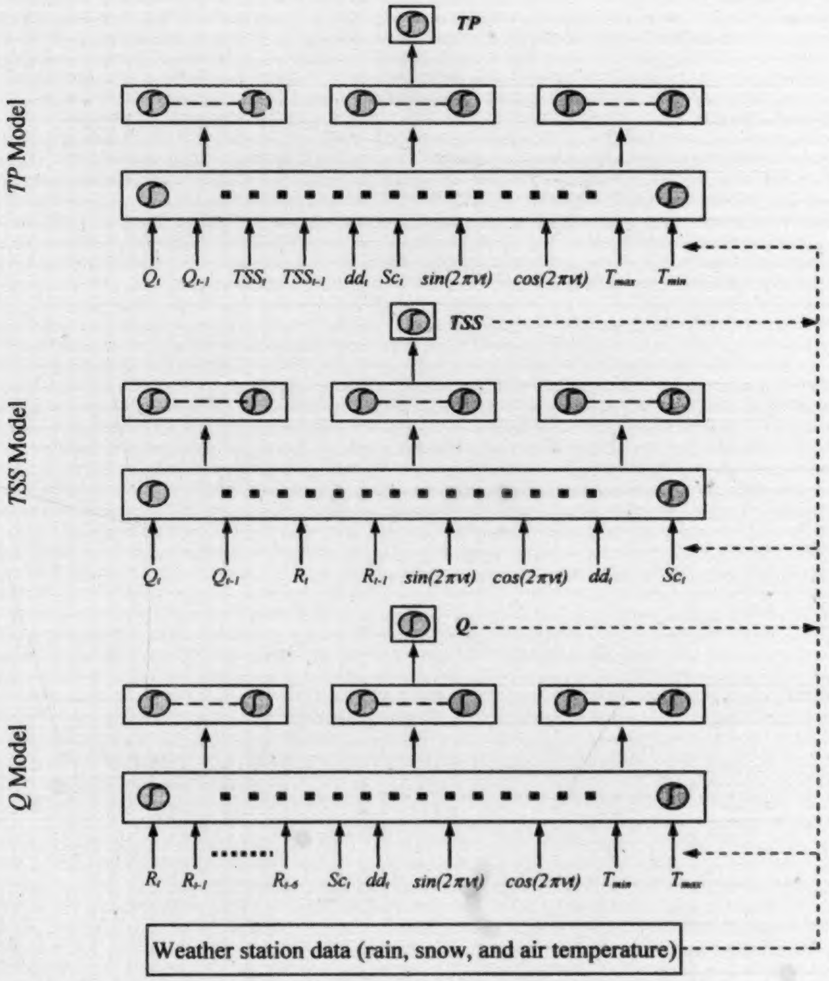

A homogeneous group of data
Kohonen layer
Inputs

Figure 6-1. A schematic showing data division using Kohonen neural networks


TP Model
TP
Q_t Q_{t-1} TSS_t TSS_{t-1} dd_t Sc_t $sin(2\pi vt)$ $cos(2\pi vt)$ T_{max} T_{min}
TSS Model
TSS
Q_t Q_{t-1} R_t R_{t-1} $sin(2\pi vt)$ $cos(2\pi vt)$ dd_t Sc_t
Q Model
Q
R_t R_{t-1} .... R_{t-6} Sc_t dd_t $sin(2\pi vt)$ $cos(2\pi vt)$ T_{min} T_{max}
Weather station data (rain, snow, and air temperature)

Note:

R_t, R_{t-1}, R_{t-2}, R_{t-3} R_{t-4}, R_{t-5}, and R_{t-6} ;	rainfall in mm/d at lags 0 through 6
T_{max}, T_{mean}, and T_{min} ;	maximum, mean, and minimum air temperatures in °C, respectively
dd_t ;	cumulative degree days at time *t*
Sc_t;	cumulative snowfall at time *t* in cm
Q_t and Q_{t-1} ;	modelled flow values in m^3/s at lags 0 and 1, respectively
TSS_t, TSS_{t-1} ;	modelled *TSS* values in mg/L at lags 0 and 1, respectively
TP ;	modelled *TP* concentration in μg/L

Figure 6-2. Optimum ANN models' architectures and model inputs

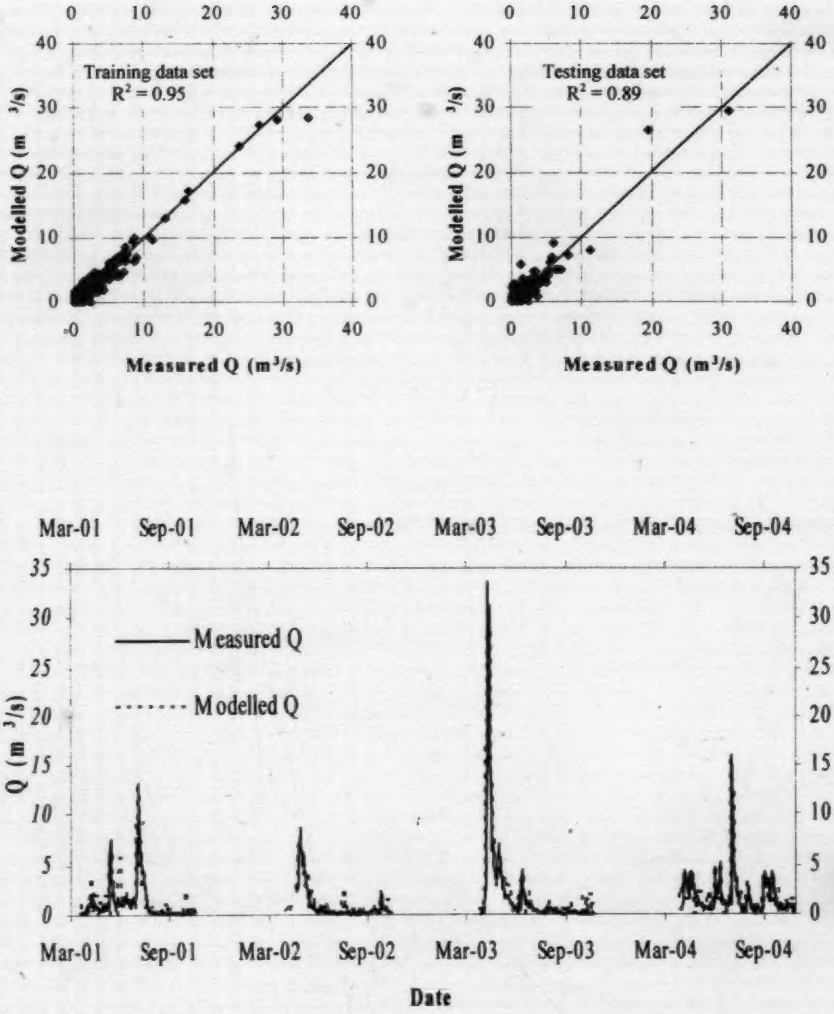


Figure 6-3. (a) Top panel; scatter plot of measured and predicted flow values and (b) bottom panel; measured vs. modelled flow hydrographs

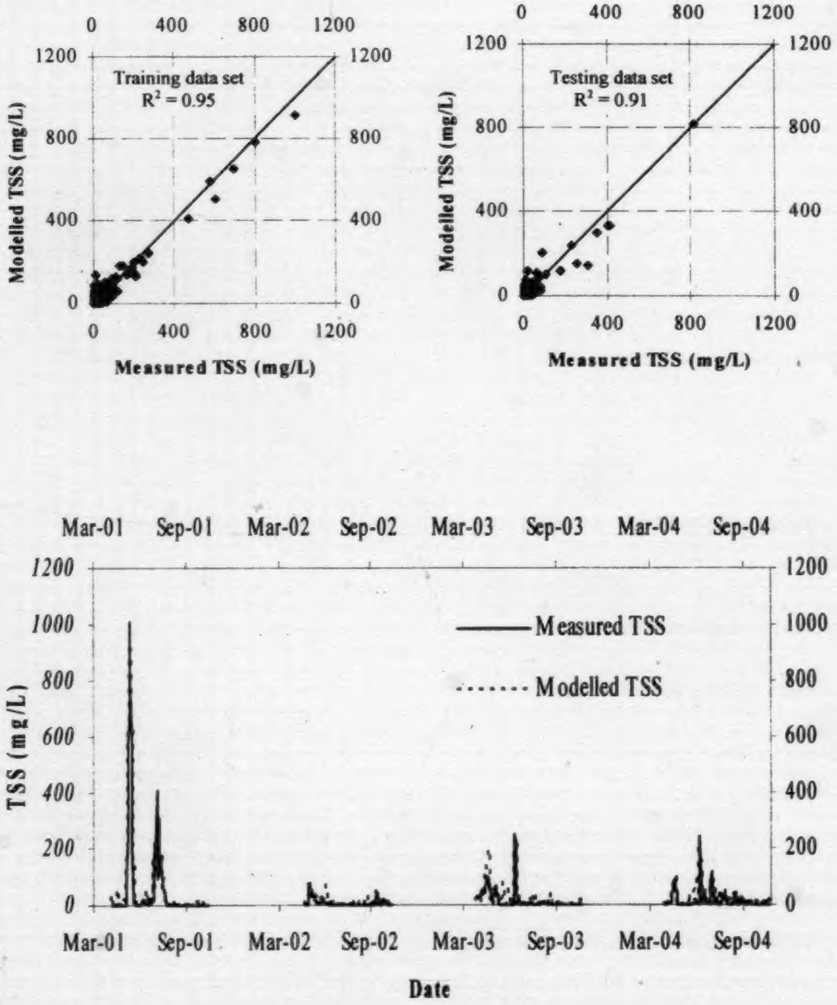


Figure 6-4. (a) Top panel; scatter plot of measured and predicted TSS values and (b) bottom panel; measured vs. modelled TSS profiles

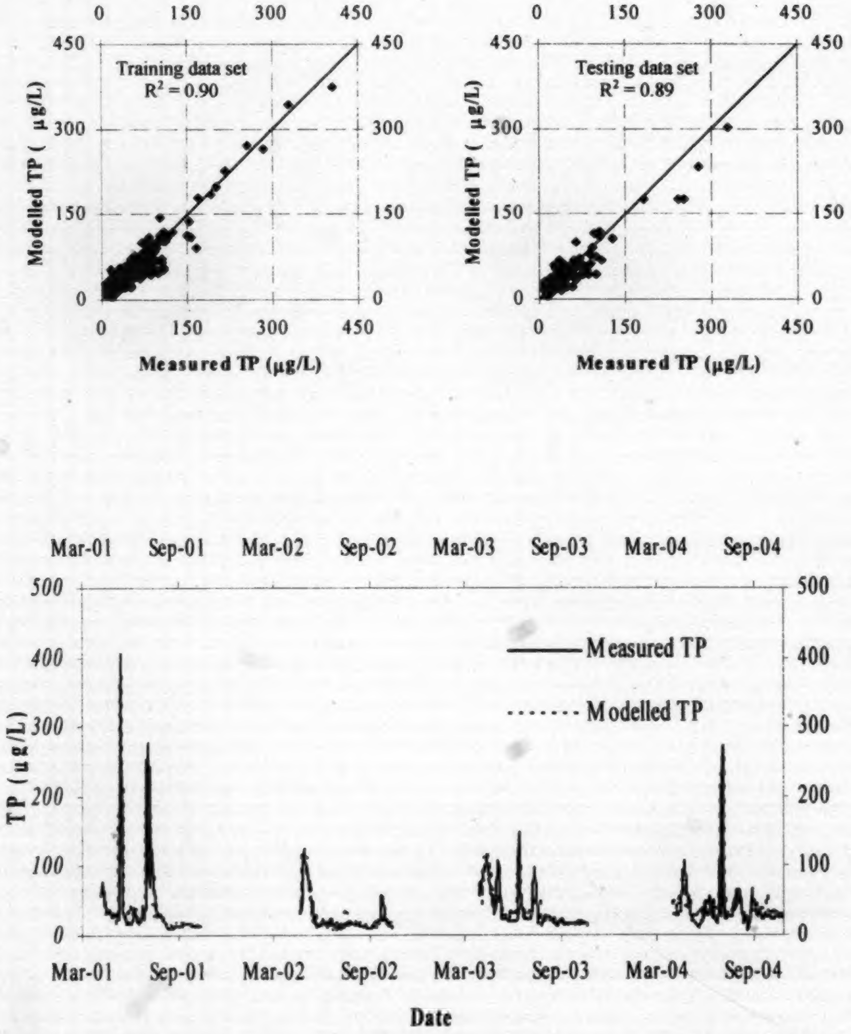


Figure 6-5. (a) Top panel; scatter plot of measured and predicted TP values and (b) bottom panel; measured vs. modelled TP profiles

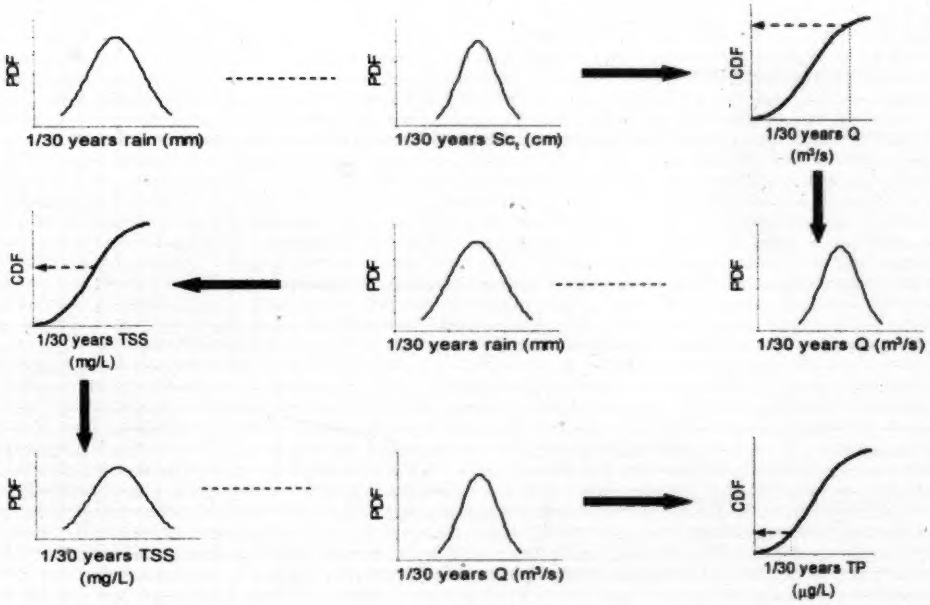

PDF
1/30 years rain (mm)
PDF
1/30 years Sc$_t$ (cm)
CDF
1/30 years Q (m^3/s)
CDF
1/30 years TSS (mg/L)
PDF
1/30 years rain (mm)
PDF
1/30 years Q (m^3/s)
PDF
1/30 years TSS (mg/L)
PDF
1/30 years Q (m^3/s)
CDF
1/30 years TP (µg/L)

Figure 6-6. A proposed framework for utilizing the developed models in water resources management

CHAPTER 7. ON THE POTENTIAL OF SATELLITE DERIVED VEGETATION INDICES FOR WATERSHED PHOSPHORUS MODELLING: A NEURAL NETWORK APPROACH

7.1. Introduction

A measurable increase in nutrient loading to water bodies may promote dissolved oxygen depletion, increased cyanobacteria biomass, and cyanobacterial toxin production. This may lead to the disruption of aquatic habitats and might also deteriorate the performance of downstream water treatment plants (Prepas et al. 2001). Therefore, the presence of cyanobateria in water bodies, used either for drinking water, animal watering, or for recreational purposes, can result in both a major economic burden and a serious health risk for human and animal populations (Hoeger et al. 2004). For example, the 1991 algal bloom of the Darling-Barwon River in Australia has been largely attributed to a high concentration of phosphorus. This bloom led to a loss of one million people-days of drinking water (Herath 1995). Moreover, an algal bloom causes downstream water treatment plants to incur additional treatment costs not only to reduce cyanobacterial cells, odor and color, but also to remove the toxins produced by the cyanobacteria. Thus, nutrient modelling, and in particular phosphorus modelling that can reflect landscape/aquatic

---

A part of this chapter has been published. Nour, M.H., Khan, A., Smith, D.W., and Gamal El-Din, M. 2005. On the potential of satellite derived vegetation phenology for watershed nutrient modelling: a neural network approach. Proceedings of the Water Environment Federation, WEFTEC® 2005. Washington, D.C. 23 pp.

phosphorus interaction, is critical in protecting aquatic ecosystems and preserving source water quality.

While it is easy to understand the hydrological, biological, and chemical mechanisms controlling phosphorus transport at the microscale, upscaling to the watershed level is not only data intensive and economically unviable, but attempts also generally yield inaccurate results (Hauhs et al. 1996; Haan 1989; Band et al. 2001; Zhao et al. 1999; Karul et al 1998; Maier et al. 1998). Instead, Artificial Neural Network (ANN) models have been relatively successful in capturing data patterns without incorporating extensive knowledge of the biological, geological, chemical, and physical processes governing the modelled system, and consequently appear to be attractive alternatives to traditional conceptual models (e.g. Brion and Lingireddy 2003; Lek et al. 1996; Maier et al. 1998; Maier and Dandy 1996; Wilson and Recknagel 2001; Moatar et al. 1999; Zhang and Stanley 1997; Zhang et al. 2004).

In the Canadian Boreal Plain, total phosphorus (TP) concentration in receiving water bodies is mostly affected by soil phosphorus content. Because of the higher susceptibility of soil to erosion during the snowmelt and storm events, it is very likely that particulate phosphorus concentration in water bodies would increase to a maximum during such events. In order to characterize export mechanisms and to predict the water-phase TP concentration in all cases, knowledge of the time patterns of both the soil nutrient concentration and the vegetation nutrient uptake is crucial. In the absence of such information, it is only possible to predict the daily change in TP concentration (Nour et al. 2006a). Providing this information in a cost effective manner is a challenge and, thus, many of the currently available models for TP predictions (both physically-based and data-driven) are undermined in practice because of the extensive landscape data required for model calibration.

Recently, rather than having to rely on a limited sampling of data, as would be the case with ground based measurements, satellite remote sensing (RS) has made cost-effective data available for the entire landscape. The Moderate-resolution Imaging Spectroradiometer (MODIS) launched by the National Aeronautics and Space Administration (NASA) in December 1999 has greatly improved scientists' ability to measure plant growth and snow cover with an adequate spatial and temporal resolution and, most importantly, at no cost to the data user. The MODIS-derived vegetation indices (VIs), the enhanced vegetation index (EVI), its predecessor, the normalized difference vegetation index (NDVI) can potentially explain phosphorus dynamics within a watershed. For example Griffith et al. (2002a, b) found significant correlations between the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) NDVI-derived phenological metrics and water quality parameters, particularly phosphorus and nitrogen. Despite this promising data, so far very few studies have focused on linking RS land use/vegetation dynamics parameters to phosphorus content in a catchment's streams. In addition, to our knowledge, no previous effort has attempted either dynamic empirical correlations of VIs and water quality parameters (as opposed to time static correlations) or has utilized such information in constructing a predictive water quality model.

This study is the first to attempt to build a model that can rely on a dynamic suite of remotely sensed vegetation indices for predicting water-phase TP concentration. It is aimed at building an ANN phosphorus model that utilizes vegetation indices, capitalizing on high quality data available via MODIS and the flexible model development procedures of the ANN technique. The specific objectives of this research were: (1) to devise a robust ANN phosphorus model, (2) to develop a protocol for linking RS data with ANN models, (3) to assess the usefulness of the MODIS-derived VIs in capturing phosphorus dynamics within a watershed, and (4) to apply the developed models to a second-order stream watershed in the Canadian Boreal forest.

7.2. Remotely Sensed Vegetation Indices

To carry out photosynthetic activities, chlorophyll in vegetation absorbs solar radiation primarily in blue and red wavelengths, permitting green to be transmitted through and reflected from the leaves. Non-visible Near Infra Red (NIR) is strongly reflected by the leaf structure to avoid over-heating. The RS vegetation indices (VIs) are derived from the manipulation of this characteristic and the unique spectral response signature of live vegetation, in order to yield a single dimensionless value that helps to distinguish plants from non-photosynthetic terrain features of the earth. While high values of VI represent dense, actively growing healthy vegetation, lower or negative values correspond to bare soil, snow, clouds or non-photosynthetic surfaces. The VI algorithms are designed to extract the vegetation signal portion from the measured reflected radiation by the remote sensor. One such algorithm is the simple ratio vegetation index (SRVI), which is the ratio of the atmospherically corrected reflectance of NIR to red wavelength. The NDVI is the normalized transform of the SRVI, designed to constrain the VI value between -1 and +1

(Tucker 1979).

All vegetation indices share similar information in regards to the relationship of the seasonal sequence of climatic factors to the timing of growth and reproductive phases in vegetation, such as the initiation of seasonal growth, time of blooming, time of seed set, and the development of new terminal buds (commonly referred to as vegetation phenology). The differences between the VIs are in the sensitivity of each indicator to the phenological cycle, the degree of saturation at high vegetation levels, and the distortion in plant response as a result of background scatter. Comparisons of the various vegetation indices can be found in Huete and Liu (1994), Elvidge and Chen (1995), Huete et al. (1997), McDonald et al. (1998) and Diaz and Blackburn (2003). Table 7-1 summarizes the vegetation indices examined as inputs to the predictive phosphorus model developed in this study.

This study examines five literature based VIs, enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), greenness fraction vegetation index (GFVI), simple ratio vegetation index (SRVI), normalized difference water index (NDWI), and two additional modified indices proposed in this study (SRVIm and GFVIm) for use as possible inputs to the model. The GFVI and SRVI were modified to be based upon EVI instead of NDVI and represented as SRVIm and GFVIm as shown in Table 7-1. Because the EVI was designed to perform better in dense vegetation (Huete et al 1994, 1999, and 2002), modified versions of SRVI and GFVI utilizing EVI in place of NDVI were proposed to cope with the forested ecosystem under study (SRVIm and GFVIm).

NDVI derived from different sensors has been used extensively by the research community in the last two decades and has been shown to correlate with several biophysical parameters such as chlorophyll density (Tucker et al. 1985), absorbed photosynthetically active radiation (Myneni and Williams 1994), leaf area index (LAI) (Spanner et al. 1990), productivity (Prince et al. 1995), and percent canopy cover (Yoder and Waring 1994). Other studies have shown the NDVI to be related to carbon-fixation (Raich and Schlesinger 1992), canopy resistance, and potential evapotranspiration (Running et al. 1989) allowing it to be used as an input to models of biogeochemical cycles (Asrar et al. 1984).

In contrast to NDVI, EVI employs the blue wavelength, which is more atmosphere-sensitive, in addition to the red and NIR, to correct the red band for aerosol influence (Huete et al. 2002). EVI is designed to provide improved vegetation monitoring in high biomass regions through the reduction of canopy background influence (Huete et al. 2002). This index is relatively new to the research community and will take time to be fully evaluated for its strengths and limitations. While the NDVI is chlorophyll sensitive, the EVI is more responsive to canopy structural variations, including LAI, canopy type, plant physiognomy, and canopy architecture (Gao et al. 2000). These two vegetation indices complement each other, though EVI is designed to be more efficient in a forested ecosystem (Huete et al. 2002).

The principle behind the derivation of the GFVI from NDVI is to relate NDVI of mixed pixels to reference NDVI values, such as the NDVI of dense vegetation ($NDVI_\infty$) and that of bare soil ($NDVI_o$), assuming the individual component NDVIs in mixed pixels can be represented by these reference NDVIs (Gutman and Ignatov 1998).

Unlike the NDVI, SRVI, GFVI, and EVI that are sensitive to the vegetation chlorophyll-content, the normalized difference water index (NDWI) is more sensitive to the change of liquid water content of vegetation canopies (Gao 1996). The NDWI relies on two spectral regions; the NIR and the mid infrared (MIR) or shortwave infrared. The inclusion of information from the MIR spectral region allows the monitoring of leaf water content (Grant 1987).

7.3. Study Area and Data Acquisition

7.3.1. Study Area and Ground-Based Data Acquisition

The study area is located in the Virginia Hills, Alberta, Canada (Figure 7-1). The studied watershed, The Willow Creek watershed, has a basin area of about 16 km^2. The area is exemplified by low topographic relief and alkaline phosphorus-rich soils (mainly fine-textured Luvisols) developed from sedimentary bedrock. The studied watershed is mostly forested. Its forests contain white spruce (Picea glauca), lodgepole pine (Pinus contorta), trembling aspen (Populus tremuloides), and balsam poplar (P. balsamifera). The climate is cool-temperate (the mean monthly air temperatures range from –23 to 18 °C), and the mean annual precipitation is 584 mm (1972 to 1997 (Environment Canada 2002)).

Mean, minimum, and maximum daily temperature data were obtained from Environment Canada's Whitecourt airport weather station due to its proximity to the studied watershed (Figure 7-1). The mean daily stream flow (Q) and TP concentration data were collected for three successive years (2001 to 2003) as part of the Forest Watershed and Riparian Disturbance (FORWARD) Study (see *J. Environ. Eng. Sci.* special issue Volume 2, 2003 for details). The meteorological data indicated that, while the hydrograph of the year 2001 was mainly influenced by storm events, snowmelt events in the years 2002 and 2003 were the most significant contributors to their hydrographs. Both the Q and TP concentration data time series peaked at almost the same time and had proportional values over time, indicating a high correlation between these variables.

7.3.2. Remote Sensing Data Acquisition

The U.S. National Aeronautics and Space Administration (NASA) has several moderate and coarse spatial resolution sensing systems in orbit that scan the entire surface of the earth and collect data. MODIS (or Moderate Resolution Imaging Spectroradiometer) is the key sensor on board NASA's Terra and Aqua satellites, which are part of the Earth Observing System (EOS) satellite constellation. The EOS program, since its creation in 1958, has focused on understanding the Earth's air, land, water, and life as an integrated system, generating an extensive long-term database of remotely sensed observations. MODIS' design is built on the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and Landsat Thematic Mapper (TM) experiences, to provide improved monitoring for land, ocean and atmospheric research. Compared to other global coverage moderate resolution spectrordiometers, MODIS has the highest spectral resolution and provides better cloud and atmospheric characterization (Justice et al. 1998). While viewing the entire earth's surface every 1 to 2 days and acquiring data in 36 spectral bands, MODIS has improved spatial resolutions of 250 m and 500 m in addition to 1 km.

The dataset is obtained in Hierarchical Data Format - Earth Observing System (HDF-

EOS), which is the standard archive format for EOS Data Information System (EOSDIS) products. HDF-EOS is a multi-object file format and supports a variety of data types. The name of the data set is "MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V004", or "MOD13Q1". L3 in the dataset name stands for Level 3, whereas collection 4 (V004) refers to reprocessed collection 1 and 3 MODLAND products that were sensed from November 2000 to date, applying the latest available version of the science algorithm and using the best calibration and geolocation information available (Land Data Operational Product Evaluation (LDOPE) 2002). The MOD13Q1 HDF-EOS file size is approximately 500 MB and consists of 11 Science data sets (SDSs), which are the actual data stored in array format (MODIS Land Science Team (MLST) 2004)). The first two layers of the MOD13Q1 file are the EVI and NDVI images. These are 16-day composite, re-sampled, 250 m spatial resolution, 4800 x 4800 rows/columns, cloud-free, pre-processed high quality imagery VI pixels, produced for each year since 2000.

Although the valid range of NDVI or EVI is from -0.2 to +1, the values are scaled up by a factor of 10,000 with a fill value of -3000 if no data was available. The next two layers provide the NDVI and EVI per-pixel quality information followed by the four SDSs of reflectance data from which NDVI and EVI are derived. This provides users with the flexibility to change the coefficient values in the EVI equation or modify the algorithms in relevance to regional conditions. The last three SDSs provide the sun-canopy-sensor angles. This information is needed in order to know the variable scan geometry under which the pixel reflectances were measured by the MODIS sensor. Each SDS (or layer) is a tile unit (fixed-area size) in a Sinusoidal (SIN) grid projection. The tile unit is the smallest unit of

MODIS land data processed at any time and has an aerial extent of approximately 1200 km x 1200 km (10^o x 10^o). The tiles are defined in a global non-overlapping grid such that there are 460 tiles, of which 326 contain land pixels.

A sequence of 3 years, from 2001 to 2003, of MODIS VIs (NDVI and EVI) and spectral reflectance (red (band 1, 620 to 670 nm); NIR (band 2, 841 to 876 nm); blue (band 3, 459 to 479 nm); and MIR (band 7, 2105 to 2155 nm)) datasets was ordered through EOS data gateway interface, accessed using MODIS' website: http://modis.gsfc.nasa.gov/ through the "data" link. The Terra MODIS MOD13Q1 dataset was chosen in the search criteria and the coordinates of the study area were fed into the system. 23 EOS-HDF files for each year; 2001, 2002 and 2003, (total of 69 files with 69 metadata files) were downloaded from NASA's server. After quality checks, 69 MOD13Q1 files were imported into ERDAS Imagine® GIS software. All of the 11 Scientific Data Sets (SDSs) produced by MODIS in each HDF-EOS file were converted to 11 image files (.img), the ERDAS Imagine® standard raster file format. The images corresponding to each spectral band of interest for each year were then stacked using the "Image Stack" module and subsetted using the Area of Interest (AOI) shape file of the study watershed. Finally, each SDS was averaged over the watershed area to present the overall watershed response with respect to the vegetation dynamics.

7.4. Vegetation Dynamics and Linkage to TP Modelling

Watershed level vegetation dynamics were captured using MODIS atmospherically corrected reflectance data and subsequently manipulated to yield various vegetation indices.

Figures 7-2 through 7-4 depict the vegetation dynamics pattern, as represented by NDVI, EVI, NDWI, GFVI, GFVIm, SRVI, and SRVIm for the Willow Creek watershed from May 2001 to October 2003. These curves conform to a typical forest phenological cycle, which has the same basic elements: from null in winter (or low) to full photosynthetic status in late spring and back to senescent in the fall. These profiles suggest that all vegetation indices share information about the relationship of the seasonal sequence of climatic factors to the timing of growth and reproductive phases in vegetation representing phenological cycles. The differences are in the sensitivity of each indicator to the phenological cycle, the degree of saturation at higher photosynthetic activity levels, and the distortion in plant response as a result of the background scatter.

NDVI based indices clearly show extreme fluctuations and appear to be very sensitive to a certain degree of photosynthetic activity in the ecosystem while EVI based indices are less sensitive to the same degree of activity. Reduced sensitivity enables EVI based indices to avoid saturation at higher levels of photosynthetic activity, indicating their potential utility in a primarily forested ecosystem such as the Willow Creek watershed.

GFVI and GFVIm represent another normalization process, constraining the values of both EVI and NDVI between 0 and 1 (for dense vegetation environments). Essentially, it stretches the values between these two extremes. Thus, both indicators are very alike in pattern and magnitude (Figure 7-4).

The NDWI behavior was different from the previous chlorophyll-based indices. The winter season with snow present, yielded the highest NDWI values. This is a consequence of NDWI being sensitive to vegetation moisture content as well as soil surface moisture or snow content. However, the typical forest phenological cycle was still captured by the NDWI (Figure 7-2).

Because of differing sensitivity and sensing patterns of each index to the same levels of photosynthetic activity, background contamination, and plant health (as a responsive indicator of soil nutrient availability and other growth limiting factors), we examined the possibility of using all seven VIs to provide landscape information for water-phase TP modelling. In addition, we conducted a sensitivity analysis designed to address the uncertainty of the studied VIs; and finally, we made recommendations on the most useful indicators in TP modelling within a forested landscape.

7.5. Methodology

This study assessed the usefulness of the MODIS-derived VIs in devising a robust ANN model for predicting the dynamics of water-phase TP concentration. The following steps summarize our methodology towards these objectives:

(1) A conceptual design for the development of parsimonious ANN TP models with a protocol for linking RS VIs to ANN was first established.

(2) Modelling TP in ungauged watersheds requires utilizing modeled flow values as inputs. However, to reduce inputs uncertainty associated with non VI inputs in this case, models employed measured flow values in place of model predicted ones. Therefore, different ANN TP models were developed in which all models utilized the same inputs except for the input representing the VI yielding seven TP models, namely:

ANNTP(EVI), ANNTP(NDVI), ANNTP(GFVI), ANNTP(GFVIm), ANNTP(SRVI), ANNTP(SRVIm), and TP(NDWI).

(3) Sensitivity analysis was performed to test the impact of the VI uncertainty on model predictions.

(4) The best performing model (based on model evaluation criteria and sensitivity analysis) was then used to predict three years of data of a ten-fold bigger watershed, the Two-Creek watershed (130 km^2) (see Figure 7-1 for location) in order to test the applicability of the proposed modelling algorithm outside the present case study.

7.6. Development of Artificial Neural Networks Models

A neural network is a computational model that is inspired by the neuron cell structure of the biological nervous system. Historical data, representing the process under consideration, is fed to the neural network during the network training, thus allowing the ANN to learn the relationships between the input(s) and the output(s). A learning rule dictates how the ANN responds to the training data. The networks consist of numerous individual processing units called neurons, interconnected in a variety of structures. The most common of these is a three layer structure called the "three-layer multilayer perceptron (MLP)". These neurons are analogous to the neurons in the human brain, which are responsible for information processing. The neurons in the input layer receive input data. Each input neuron represents a single input parameter and scales inputs in a numeric range that is consistent with the training scheme. The neurons in the hidden layer process the data through a set of non-linear activation functions (transfer functions) achieving the nonlinearity of the network. The output layer neurons report the results from the network in the original numeric scale (Haykin 1994). Each neuron in a layer is connected to every neuron in the subsequent layer (feed-forward (FF) connections). These connections are similar to the human dendrites and axons that allow the communication between the neurons in the brain. In the human brain, signals are transferred between neurons through these connections and across the synaptic gap, resulting in the release of chemicals that stimulate or inhibit the ability of the neighboring neurons to generate impulses. In the ANN, the "connection weight" between neurons represents this communication process. The sign and magnitude of connection weights describe the nature and strength of influence between the connected neurons (Smith 1993).

During the learning process, ANNs apply a set of rules that change connection weights iteratively until a stopping criterion is reached. Maier and Dandy (2000) and Zealand et al. (1999) reported that over 80% of previous neural network models used a backpropagation (BP) training algorithm, which is a supervised learning paradigm. By supervised learning, it is meant that a desired response is available to guide the learning process. In the BP algorithm, the weights are initially assigned small values arbitrarily. As training progresses, the mean squared error (MSE) between the target output and the network output is calculated, and the weights are systematically updated. Weight adjustments are made based on an objective function that reduces the MSE, attempting to reach a global minimum in the error surface. The training process stops when a prescribed stopping criterion is reached.

The development of the ANN models in this study can be summarized by three important phases: data pre-processing and input determination, model building, and model evaluation.

7.6.1. Data Pre-processing and Input Determination Phase

TP concentration in receiving water bodies of the Boreal Plain is mostly affected by soil phosphorus concentration and the intersection of the water table with the surface soil layers. Thus, snow melt and storm events enhance phosphorus migration to water bodies because the soil is more susceptible to erosion during these events (Chanasyk et al. 2003). Therefore, Q (being the outcome of all hydrologic processes defining the study watershed) and remotely sensed VIs (as a representation of soil/vegetation phosphorus interaction), with proper time series manipulation, can serve as inputs to a dynamic TP concentration model.

During the data pre-processing, TP concentration and Q time series indicated an annual cyclic variation, a seasonal periodicity, and TP/Q hystereses loops. Thus, when modelling TP, not only causal inputs were included in model formulation, but also input reflecting time correlation, and the TP/Q hystereses seasonal variations. An adequate understanding of the physical processes deriving the modeled parameter is critical to identifying its causal inputs; time series analysis (cross-correlation analysis and spectral analysis) can identify time-lagged inputs and can feed the ANN model with information to indicate hystereses loops.

Efficient manipulation of cyclic time series was attained by conducting spectral analysis to the data. The variance profile over the frequency, usually referred to as the power spectrum, was constructed to identify the frequency that contributed most to the variance. The monthly variation, indicating the changes between months within the typical annual cycle, was identified as the dominant data periodicity (Figure 7-5). In reflection to that, two additional ANN model inputs, $\sin(2\pi\frac{t}{12})$ and $\cos(2\pi\frac{t}{12})$ were introduced to the model in order to account for the cyclic nature of the studied series. Figure 7-6 illustrates this approach. By detecting the sign of the two inputs (either positive or negative), the model is thought to be able to identify the season under study (e.g., a positive value for both the sine and cosine inputs identifies the winter season), and with the aid of the magnitude of these inputs, the model is expected to be able to identify the month within each season. The inclusion of these two additional inputs likely enabled the model to dynamically change input/output relation according to the season and, thus, to address TP/Q hystereses.

The cross-correlation analysis was used to address inputs' time correlation and to identify possible time-lagged inputs based on the strength of the correlation between the output variable and each of the time-lagged inputs. It was found that TP is correlated to Q(t), Q(t-1), Q(t-2), and Q(t-3). A Detailed description of the spectral analysis and the cross-correlation analysis is presented in Nour et al. (2006b).

Seven models were constructed in this study; all models utilized the following inputs as dictated by input/output causality and time series manipulation: Q(t), Q(t-1), Q(t-2), Q(t-3), $\sin(2\pi\frac{t}{12})$, $\cos(2\pi\frac{t}{12})$, and the mean average daily air temperature (T_{avg}). In addition to the previous seven inputs, each of the seven developed models utilized a different VI (EVI, NDVI, SRVI, SRVIm, GFVI, GFVIm, and NDWI).

7.6.2. Model Building Phase

The goal of this phase is to build a parsimonious model that is not limited to mapping the data used during its development but that can also generalize by mapping other data sets. Thus, the choice of the training algorithm, the network architecture and internal parameters (number of hidden layers, number of neurons, the type of scaling and activation functions, the learning and momentum rates, and the stopping criterion), and the division of data into training (calibration), testing (validation), and cross-validation data sets, are crucial to achieving a robust model. A thorough description of the model-building phase is presented in Nour et al. (2006b), and Maier and Dandy (2000). Following their guidelines in building a robust model, a three-layered feedforward multilayer perceptron (FF-MLP) ANN model, trained with the error backpropagation algorithm, was found to perform best for this case study. In all cases, the optimum model employed a single hidden layer with three different activation functions (it appears that the three activation functions are analogous to the three driving forces in the modeled system: base flow, snow melt, and storm events.) Figure 7-7 portrays the structure of the ANN model that was optimum in all the seven devised models.

Two training algorithms were tested in this study: firstly, the typical gradient descent BP algorithm that utilizes a learning rate and a momentum coefficient to control the training speed and to facilitate moving towards a global minimum in the error surface

(Haykin 1994), and secondly, a BP algorithm with a batch update (BP-BU) technique. The NeuroShell 2® software package was used to train both models (Ward Systems 1996). In the batch mode of the BP learning, training proceeds through an entire epoch (i.e. it cycles through all of the training data set patterns) before the weights are updated. The advantage of the second algorithm (the BP-BU algorithm) is that it is insensitive to both the learning rate and the momentum coefficient, giving flexibility to less experienced modelers (Gamal El-Din and Smith 2002).

The data was divided into three portions in the ratio of 3:1:1 for training (T), testing (S), and cross validating (V) the model, respectively. The training data set was used to adjust the connection weights. The testing data set determined the point at which training has to be stopped to prevent overfitting the data. The ability of the model to generalize (i.e. to produce correct results on previously unseen data, rather than just to memorize the data already encountered during training) is finally measured by applying the developed model to the cross-validation data set. Table 7-2 summarizes the features of the optimum ANN model in each of the devised models.

7.6.3. Model Evaluation Phase

During the model evaluation phase, the coefficient of multiple determination (R^2) was used as a statistical measure of goodness of model fit, along with the root mean squared error (RMSE). Measured versus predicted profiles were graphically examined, and the model residuals were analyzed and checked for possible trends not explained by the developed model. In addition, in order to test the robustness of each candidate model, the testing (S) and the cross-validation (V) data sets were swapped, each model was retrained, and the new model performances were then monitored.

7.7. Results and Discussion

7.7.1. Modelling Results

Following the model development algorithm and the methodology section described above, seven ANN models were built and applied to the Willow Creek watershed. Table 7-2 illustrates the ANN model configuration for each of the developed models. The robustness of the modelling approach in terms of the model architecture is portrayed by the similarity in the model configurations among all the devised models. The only difference found was in the number of hidden neurons. The statistical performances of all models are summarized in Table 7-3. The devised ANN TP models for the Willow Creek watershed were able to simulate TP concentration successfully. All EVI-based VIs performed better than their corresponding NDVI-based VIs with R^2 ranges of (0.74 to 0.98) and (0.63 to 0.96) for the EVI-based and NDVI-based models, respectively. This result can be supported by the fact that the watershed under study is a primarily forested watershed. The EVI was designed to serve densely vegetated ecosystems by preventing the index saturation at high photosynthetic activity levels and consequently performed well when used as an input for modelling TP concentration. ANNTP(EVI) and ANNTP(GFVIm) models performed almost identically in their results because in ANN, the data is internally normalized through a linear scaling function analogous to the transformation of EVI to GFVIm and thus would deal with these indicators in almost the same way. The same trend was detected for their corresponding NDVI-based models: ANNTP(NDVI) and ANNTP(GFVI). The minor variation seen in their performance; however, might be attributed to the difference in the normalization range between the ANN scaling function and the function that transforms NDVI and EVI to GFVI and GFVIm, respectively. The ANNTP(NDWI) model performed slightly better than the ANNTP(NDVI) but worse than the ANNTP(EVI). SRVI and SRVIm were the poorest VIs in representing water-phase TP dynamics.

The ANN models developed proved to be stable and consistent in their prediction. This was reflected by the maintained high R^2 value, even when the testing and cross-validation data sets were swapped. The maximum root mean squared error (RMSE) for all data sets was 40 μg/L (corresponding to the SRVI). The RMSE was small compared to the magnitude of its corresponding TP concentration, highlighting the high performance of the developed models in all data cases. The residual analysis was then conducted as a final stage in the model evaluation. Graphing the residuals is very important in model judgment; if the model fit the data well, the residuals can only be expected to reflect the measurement error that is assumed to be random. Hence, any lack of randomness in the residuals undermines the strength of the fitted model. In this case, the residuals were plotted versus time and versus flow to check for residual independency. Residual plots showed only random scatter (no trends were detected, instead, points were scattered all over the graph plot), indicating that the developed models have no serious deficiencies.

The results detailed above indicated that either the EVI or its normalized transform, GFVIm, can successfully represent landscape phosphorus dynamics and can therefore serve as inputs to a TP ANN model for adequate TP predictions. Their corresponding NDVI-based indices, as well as the NDWI were slightly lower in their performances, yet they can still provide ample information for fair TP modelling, especially for periods when EVI values are missing given that NDVI has more than 20 years of historic data.

Finally, to be able to select the optimum VI in terms of representing landscape phosphorus dynamics in a TP model, sensitivity analysis was conducted to test the model robustness in response to the VIs uncertainty. The following section will summarize the conducted sensitivity analysis.

7.7.2. Sensitivity Analysis

In response to input parameters' uncertainty, quantifying the uncertainties in model predictions is important in evaluating a modelling tool. Although a robust model is supposed to be largely insensitive to expected errors in the input VI, it has to be sensitive enough to capture changes responsive to land use and watershed management activities. A sensitivity analysis is presented here to evaluate the robustness of all candidate models with respect to the impact of VIs uncertainty on TP models' predictions skill.

The MODIS land science team has evaluated VI accuracies in response to the reflectance calibration uncertainty, possible spectral band shift, and band-to-band coregistration error. Their calculations have revealed an estimated error in the range of ± 3% VI units for NDVI and ± 4% VI units for EVI in cloud free conditions (MLST 2004). MODIS has an algorithm designed to eliminate cloud contamination. Thus, ± 5% VI units will be used as the threshold to assess the model robustness. The VIs were also allowed to vary up to ± 10% in order to evaluate the vigor of the sensitivity, in terms of the responsive changes in TP concentration as a result of changes in the landscape vegetation composition. Figures 7-8 through 7-14 depict the sensitivity scatter plots for all the developed ANN models.

In terms of a ± 5% change in the VI level, the EVI-based models and NDWI-based model were found to be more robust when compared to the NDVI-based models. This observation is reflected in retaining a high R^2 value by the ANNTP(EVI) model (Figure 7-8), the ANNTP(GFVIm) model (Figure 7-11), and the ANNTP(NDWI) model (Figure 7-14). The minimum R^2 obtained in response to a ± 5% change in VI was 0.88 (only 6% lower than the original value) for the EVI-based models, and was 0.85 (only 6.5% lower than the original value) for the NDWI-based model. The ANNTP(NDVI) model was slightly lower in performance (8 % decrease in R^2.) Surprisingly, the ANNTP(GFVI) model was not sensitive at all to changes in the GFVI values. ANNTP(SRVI) and ANNTP(SRVIm) models were, as previously observed, inferior in performance.

In terms of the ability of the model to capture variation in vegetation cover and to predict the impact of such variation on the water-phase TP concentration (measured here by the ± 10% VI change), ANNTP(EVI) and ANNTP(GFVIm) remain superior, as the impact of the ± 10% change in vegetation typically resulted in a similar variation as predicted for a forested watershed of such scale (Prepas et al. 2001). Whereas ANNTP(NDVI) was very sensitive to the 10% change in NDVI (R^2 dropped to 0.5), ANNTP(GFVI) was insensitive to this level of disturbance suggesting that the NDVI-based VIs are not the best alternatives for a forested landscape. The ANNTP(NDWI) model dropped in performance (in terms of

R^2) from 0.91 to 0.74 by decreasing the value of the NDWI by 10%.

Based on the models' performances and sensitivity analyses, the ANNTP(EVI) was selected as the best candidate model for modelling TP concentration in a forested watershed. To test the applicability of the approach outside the presented case study, the same modelling algorithm was applied to another watershed in the Canadian Boreal Plain as described in the following section.

7.8. Model Applications

Assuming that we do not have either flow or TP measurements in the watershed of interest and that the ungauged watershed behaves as the Willow Creek watershed, one would be able to obtain real time forecasting of TP concentration by utilizing the developed ANNTP(EVI) model as follows: (1) download MODIS derived EVI as pixel based EVI values; (2) average the pixel based values over the entire watershed; (3) simulate the flow values by using a hydrologic model capable of modelling ungauged watersheds (eg., the hydrologic model developed by Nour et al. (2006a); and (4) estimate the time-lagged inputs and the inputs reflecting TP/Q hystereses as discussed in the "Data Pre-processing and Input Determination Phase" section. At such point, all model inputs can be available to run the ANN model for TP dynamics prediction.

To test the applicability of the modelling approach outside the current case study, the above algorithm was used to model TP concentrations in the Two-Creek watershed (Figure 7-1). The inputs to the model were: Q(t), Q(t-1), Q(t-2), $\sin(2\pi \frac{t}{12})$, $\cos(2\pi \frac{t}{12})$, EVI, and the mean average daily air temperature (T_{avg}). The optimum ANN architecture was found to be similar to the previous ANN models in architecture; a three-layer MLP with one hidden layer that contains 21 neurons (7 using the Gaussian, 7 using the Gaussian Complement, and 7 using the logistic activation functions). The input scaling function was the linear function and the output was the logistic activation function. The developed model performance in terms of (R^2, RMSE in μg/L) was (0.94, 9), (0.88, 14), and (0.86, 16) for the training, the testing, and the cross-validation data sets, respectively. The high model performance was maintained when the testing and the cross-validation were swapped yielding (R^2, RMSE) of (0.95, 9), (0.88, 15), and (0.85, 16) for the training, the testing, and the cross-validation data sets, respectively. The measured and the predicted TP concentrations profiles were in good agreement highlighting the success of model in predicting the TP concentration dynamics in the Two-Creek watershed (Figure 7-15).

7.9. Conclusions and Recommendations

To our knowledge, this study is the first to attempt to build a model that relies on a time series of remotely sensed vegetation indices for predicting the dynamics of water-phase TP concentration. Because of the different degrees of sensitivity of each vegetation index (VI) to ecosystem disruption and plant health (as a responsive indicator of soil nutrient availability and other growth-limiting factors), we examined the possibility of using seven VIs (NDVI, EVI, NDWI, GFVI, GFVIm, SRVI, and SRVIm) to provide sufficient landscape information for water-phase TP modelling. In addition, we conducted sensitivity analyses to quantify the uncertainty in model predictions in response to the uncertainty in the VIs. As a result, a protocol for constructing parsimonious ANN TP models and for linking RS VIs to ANN was developed. Seven ANN TP models were devised in this study: ANNTP(EVI), ANNTP(NDVI), ANNTP(NDWI), ANNTP(GFVI), ANNTP(GFVIm), ANNTP(SRVI), and ANNTP(SRVIm). All models utilized the same information except for the input representing the VI. The EVI-based models were found to be superior in terms of the model prediction accuracy, the model robustness and stability. The R^2 values for the EVI-based models ranged from 0.8 to 0.98 for all data sets, and the results of the sensitivity analyses confirmed the robustness of the devised models to possible errors in the EVI values, as well as its sensitivity to vegetation disruption in response to watershed disturbance practices.

To test the applicability of the modelling approach outside the current case study, the proposed modelling algorithm was used to model TP concentrations in another watershed, the Two-Creek watershed (130 km^2). The predicted TP concentration profile matched the measured one very well. The high model performance was further confirmed by the consistent high values of R^2 for all data cases, reflecting the superiority of the presented modelling approach.

This improved model responses over previous TP models can be attributed to the incorporation of the MODIS VI profiles, particularly the EVI in this case, in the modelling process. This is because identifying low and peak photosynthetic levels in the growing season has considerable implications for nutrient uptake and export studies. Thus, the present study serves as a leading effort in incorporating low-cost time-variant information in constructing a predictive tool for the water-phase TP concentration. It can potentially serve as a valuable tool in simulating the impact of different watershed harvesting activities on water quality parameters.

The implications of phenological information on hydrological and nutrient modelling, however, have not yet been completely explored in full depth. Thus, further studies are required to confirm the applicability of this approach for different landscape dominated watersheds. In particular, low density vegetated watersheds should be examined as NDVI-based models may perform better.

7.10. References

Asrar, G., Fuchs, M., Kanemasu, E.T., and Hatfield, J.L. 1984. Estimating absorbed photosynthetic radiation and leaf-area index from spectral reflectance in wheat. Agron. J. **76**: 300-306.

Band, L.E., Tague, C.L., Groffman, P., and Belt, K. 2001. Forest ecosystem processes at the watershed scale: hydrological and ecological controls of nitrogen export. Hydrol. Processes **15**: 2013-2028.

Brion, G.M., and Lingireddy, S. 2003. Artificial neural network modelling: a summary of successful applications relative to microbial water quality. Water Sci. Technol. **47**(3): 235-240.

Chanasyk, D.S., Whitson, I.R., Mapfumo, E., Burke, J.M., and Prepas, E.E. 2003. The impacts of forest harvest and wildfire on soils and hydrology in temperate forests: A baseline to develop hypotheses for the Boreal Plain. J. Environ. Eng. Sci. **2**: S51-S62.

Diaz, B.M., and Blackburn, G.A. 2003. Remote sensing of mangrove biophysical properties: evidence from a laboratory simulation of the possible effects of background variation on spectral vegetation indices. Int. J. Remote Sens. **24**: 53-73.

Elvidge, C.D., and Chen, Z.K. 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indexes. Remote Sens. Environ. **54**: 38-48.

Gamal El-Din, A.G. and Smith, D.W. 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. **36**(5): 1115-1126.

Gao, B.C. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. **58**: 257-266.

Gao, X., Huete, A.R., Ni, W.G., and Miura, T. 2000. Optical-biophysical relationships of vegetation spectra without background contamination. Remote Sens. Environ. **74**: 609-620.

Grant, L. 1987. Diffuse and specular characteristics of leaf reflectance. Remote Sens. Environ. **22**: 309-322.

Griffith, J.A., Martinko, E.A., Whistler, J.L., and Price, K.P. 2002a. Interrelationships among landscapes, NDVI, and stream water quality in the US central plains. Ecol. Appl. **12**: 1702-1718.

Griffith, J.A., Martinko, E.A., Whistler, J.L., and Price, K.P. 2002b. Preliminary comparison of landscape pattern-normalized difference vegetation index (NDVI) relationships to central plains stream conditions. Journal Environ. Qual. **31**: 846-859.

Gutman, G., and Ignatov, A. 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. Int. J. Remote Sens. **19**: 1533-1543.

Haan, C.T. 1989. Parametric in hydrologic modelling. T ASAE. **32**(1): 137-146.

Hauhs, M., Neal, C., Hooper, R., and Christophersen, N. 1996. Summary of a workshop on ecosystem modelling: The end of an era?. Sci. Total Environ. **183**: 1–5.

Haykin, S. 1994. Neural Networks: A Comprehensive Foundation. 2nd. ed. Macmillian College Publishing, Inc., NY, 842 pp.

Herath, G. 1995. The algal bloom problem in Australian waterways: an economic appraisal. Rev. Marketing Agr. Econ. **63**: 77-86.

Hoeger, S.J., Shaw, G., Hitzfeld, B.C., and Dietrich, D.R. 2004. Occurrence and elimination of cyanobacterial toxins in two Australian drinking water treatment plants. Toxicon. **43**: 639-649.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. **83**: 195-213.

Huete, A., Justice, C., and Leeuwen, W.V. 1999. MODIS Vegetation Index (MOD 13). Algorithm Theoretical Basis Document, Version 3. Available from http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf. [accessed on 11 September 2004].

Huete, A., Justice, C., and Liu, H. 1994. Development of vegetation and soil indexes for modis-eos. Remote Sens. Environ. **49**: 224-234.

Huete, A.R., and Liu, H.Q. 1994. An error and sensitivity analysis of the atmospheric-correcting and soil-correcting variants of the ndvi for the modis-eos. IEEE T. Geosci. Remote. **32**: 897-905.

Huete, A.R., Liu, H.Q., Batchily, K., and vanLeeuwen, W. 1997. A comparison of vegetation indices global set of TM images for EOS-MODIS. Remote Sens. Environ. **59**: 440-451.

Jordan, C.F. 1969. Derivation of leaf area index from quality of light on the forest floor. Ecology. **50**(4): 663.

Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Wan, Z.M., Huete, A.R., van Leeuwen, W., Wolfe, R.E., Giglio, L., Muller, J.P., Lewis, P., and Barnsley, M.J. 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. IEEE T. Geosci. Remote. **36**: 1228-1249.

Karul, C., Soyupak, S., and Germen, E. 1998. A new approach to mathematical water quality modeling in reservoirs: Neural networks. Int. Rev. Hydrobiol. **83**: 689-696.

Land Data Operational Product Evaluation (LDOPE). 2002. What is a product collection?; MODIS Land Quality Assessment. http://landweb.nascom.nasa.gov/cgi-bin/QA_WWW/ newPage.cgifileName=collection.html. [accessed on 6 September 2004].

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus concentration and phosphorus load from watershed characteristics using backpropagation neural networks. Acta Oecol. **17**(1): 43-53.

Maier, H.R., and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. **32**(4): 1013-1022.

Maier, H.R., Dandy, G.C., and Burch, M.D. 1998. Use of artificial neural networks for modelling cyanobacteria Anabaena spp. in River Murray, South Australia. Ecol. Model. **105**: 257-272.

Maier, H.R. and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw.**15**: 101-124.

McDonald, A.J., Gemmell, F.M., and Lewis, P.E. 1998. Investigation of the utility of spectral vegetation indices for determining information on coniferous forests. Remote Sens. Environ. **66**: 250-272.

Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural networks. Application of control and validation data series in Middle Loire River. Ecol. Model. **120**: 141-156.

MODIS Land Science Team (MLST). 2004. MODLAND Product User Guides; MODIS Land Quality Assessment. Available from http://landweb.nascom.nasa.gov/cgibin/QA_WWW/newPage.cgifileName=modland_guide. [accessed on 26 August 2004].

Myneni, R.B., and Williams, D.L. 1994. On the relationship of FAPAR and NDVI. Remote Sens. Environ. **49**: 200-211.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006a. The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. Ecol. Model. **191**: 19-32.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006b. Artificial neural networks and time series modelling of TP concentration in boreal streams: A comparative approach. J. Environ. Eng. Sci. **5**: S39-S52.

Prepas, E.E., Pinel-Alloul, B., Planas, D., Méthot, G., Paquet, S., and Reedyk, S. 2001. Forest harvest impacts on water quality and aquatic biota on the Boreal Plain: introduction to the TROLS lake program. Can. J. Fish. Aquatic Sci. **58**: 421-436.

Prince, S.D., Goetz, S.J., and Goward, S.N. 1995. Monitoring primary production from earth observing satellites. Water Air Soil Pollution. **82**: 509-522.

Raich, J.W., and Schlesinger, W.H. 1992. The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. Tellus. **44**: 81-99.

Running, S.W., Nemani, R.R., Peterson, D.L., Band, L.E., Potts, D.F., Pierce, L.L., and Spanner, M.A. 1989. Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. Ecology. **70**; 1090-1101.

Smith, M. 1993. Neural Networks for Statistical Modelling. ITP Publishing, New York, 235 pp.

Spanner, M.A., Pierce, L.L., Running, S.W., Peterson, D.L. 1990. The seasonality of AVHRR data of temperate coniferous forests: Relationship with leaf area index. Remote Sens. Environ. **33**: 97-112.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. **8**: 127-150.

Tucker, C.J., Townshend, J.R.G., and Goff, T.E. 1985. African land-cover classification using satellite data. Science. **227**: 369-375.

Ward Systems. 1996. NeuroShell 2® version 3, reference manual; Ward Systems Group, Inc., Frederick, Maryland, USA.

Wilson, H., and Recknagel, F. 2001. Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. Ecol. Model. **146**: 69-84.

Yoder, B.J., and Waring, R.H. 1994. The normalized difference vegetation index of small Douglas-Fir canopies with varying chlorophyll concentrations. Remote Sens. Environ. **49**: 81-91.

Zealand, C.M., Burn, D.H., and Simonovic, S.P. 1999. Short term stream flow forecasting using artificial neural networks. J. Hydrol. **214**: 32-48.

Zhang, Q., and Stanley, S.J. 1997. Forecasting raw-water quality parameters for North Saskatchewan River by neural network modelling. Water Res. **31**: 2340-2350.

Zhang, Q.J., Cudrak, A.A., Shariff, R., and Stanley, S.J. 2004. Implementing artificial neural network models for real-time water colour forecasting in a water treatment plant. J. Environ. Eng. Sci. **3**: S15-S23.

Zhao, H., Hao, O.J., and McAvoy, T. J. 1999 Approaches to modelling nutrient dynamics: ASM2, simplified model and neural nets. Water Sci. Technol. **39**(1): 227-234.

Table 7-1. A summary table of the vegetation indices explored in this study

Acronym	VI Description	Formula	Reference
NDVI	Normalized difference vegetation index	$\mathrm{NDVI} = \frac{\lambda_{NIR} - \lambda_{red}}{\lambda_{NIR} + \lambda_{red}}$	Tucker (1979)
EVI	Enhanced vegetation index	$\mathrm{EVI} = G\frac{\lambda_{NIR} - \lambda_{red}}{\lambda_{NIR} + C_1 \times \lambda_{red} - C_2 \times \lambda_{blue} + L}$	Huete et al. (1994)
SRVI	Simple ratio vegetation index	$\mathrm{SRVI} = \frac{\lambda_{NIR}}{\lambda_{red}} = \frac{1 + NDVI}{1 - NDVI}$	Jordan (1969)
SRVIm	Modified simple ratio vegetation index	$\mathrm{SRVIm} = \frac{1 + EVI}{1 - EVI} = \frac{(G+1)\lambda_{NIR} + (C_1 - 1)\lambda_{red} - C_2\lambda_{blue} + L}{(G-1)\lambda_{NIR} - (C_1 + 1)\lambda_{red} + C_2\lambda_{blue} - L}$	This study
NDWI	Normalized difference water index	$\mathrm{NDWI} = \frac{\lambda_{NIR} - \lambda_{MIR}}{\lambda_{NIR} + \lambda_{MIR}}$	Gao (1996)
GFVI	Greenness fraction vegetation index	$\mathrm{GFVI} = \frac{NDVI - NDVI_o}{NDVI_g - NDVI_o}$ $NDVI_g = NDVI_\infty - \left(NDVI_\infty - NDVI_o\right)e^{-kL_g}$	Gutman and Ignatov (1998)
GFVIm	Modified greenness fraction vegetation index	$\mathrm{GFVI} = \frac{EVI - EVI_o}{EVI_g - EVI_o}$ $EVI_g = EVI_\infty - \left(EVI_\infty - EVI_o\right)e^{-kL_g}$	This study

where, λ_{red}, λ_{blue}, λ_{NIR}, and λ_{MIR} are the atmospherically corrected surface reflectance at the red, blue, near infra red, and mid infra red wavelengths; L is the canopy background adjustment (to correct for non-linearity, differential NIR and red radiant transfer through a canopy); C_1 and C_2 are coefficients of the aerosol resistance term, which use the blue band to correct for aerosol influences in the red band; and G is the gain factor. The coefficients adopted in the EVI algorithm are $L = 1$, $C_1 = 6$, $C_2 = 7.5$, and $G = 2.5$ (values after Huete et al. 2002). L_g is the number of vegetation layers; $NDVI_o$ and $NDVI_\infty$ are the NDVI signals corresponding to bare soil ($L_g \rightarrow 0$) and dense green vegetation ($L_g \rightarrow \infty$), respectively; and k is the extinction coefficient.

Table 7-2. Summary table showing optimum ANN models' architecture and ANN internal parameters

Model	Data division (T:S:V)	Scaling function	Training algorithm	Learning rate	Momentum coefficient	Initial weights	Stopping criterion	Optimum network (I-[H-H-H]-O)
TP*	3:1:1	Linear, <-1,1>	BP-BU	Insensitive	Insensitive	Random, [-0.3,0.3]	Best test set (in terms of MSE)	7L-[5G-5GC-5LO]-tanh
TP(NDVI)								8L-[8G-8GC-8LO]-LO
TP(EVI)								8L-[12G-12GC-12LO]-LO
TP(GFVI)								8L-[8G-8GC-8LO]-LO
TP(GFVIm)								8L-[8G-8GC-8LO]-LO
TP(SRVI)								8L-[13G-13GC-13LO]-LO
TP(SRVIm)								8L-[12G-12GC-12LO]-LO
TP(NDWI)								8L-[6G-6GC-6LO]-tanh

where: I and O denote input and output layers, respectively; [H-H-H] represents a single hidden layer with a different activation function; L is the linear scaling function; G, GC, and LO are the Gaussian, Gaussian complement, and logistic activation functions, respectively; T, S, and V are the training, the testing, and the cross-validation data sets, respectively; and < > denotes an open interval. TP* model is the background model that does not use VIs in the vector of inputs.

Table 7-3. Statistical measures of models' performances

Model	R^2, S used as testing data set			R^2, V used as testing data set			RMSE, S used as testing data set			RMSE, V used as testing data set		
	T	S	V	T	S	V	T	S	V	T	S	V
TP*	0.88	0.41	0.48	0.79	0.58	0.32	24	55	44	32	39	59
TP(NDVI)	0.81	0.78	0.71	0.91	0.77	0.76	32	29	35	21	32	30
TP(EVI)	0.98	0.88	0.82	0.98	0.81	0.88	9	21	28	9	28	21
TP(GFVI)	0.8	0.78	0.71	0.91	0.77	0.76	32	29	35	21	32	30
TP(GFVIm)	0.98	0.89	0.84	0.98	0.89	0.87	10	20	26	10	22	22
TP(SRVI)	0.96	0.79	0.61	0.89	0.72	0.63	14	28	41	24	35	37
TP(SRVIm)	0.98	0.91	0.74	0.96	0.79	0.86	9	19	34	15	30	23
TP(NDWI)	0.97	0.83	0.77	0.96	0.85	0.78	13	25	32	14	25	28

where: T is the training data set; S is the testing data set; and V is the cross-validation data set. RMSE is in (μg/L). TP* model is the background model that does not use VIs in the vector of inputs.

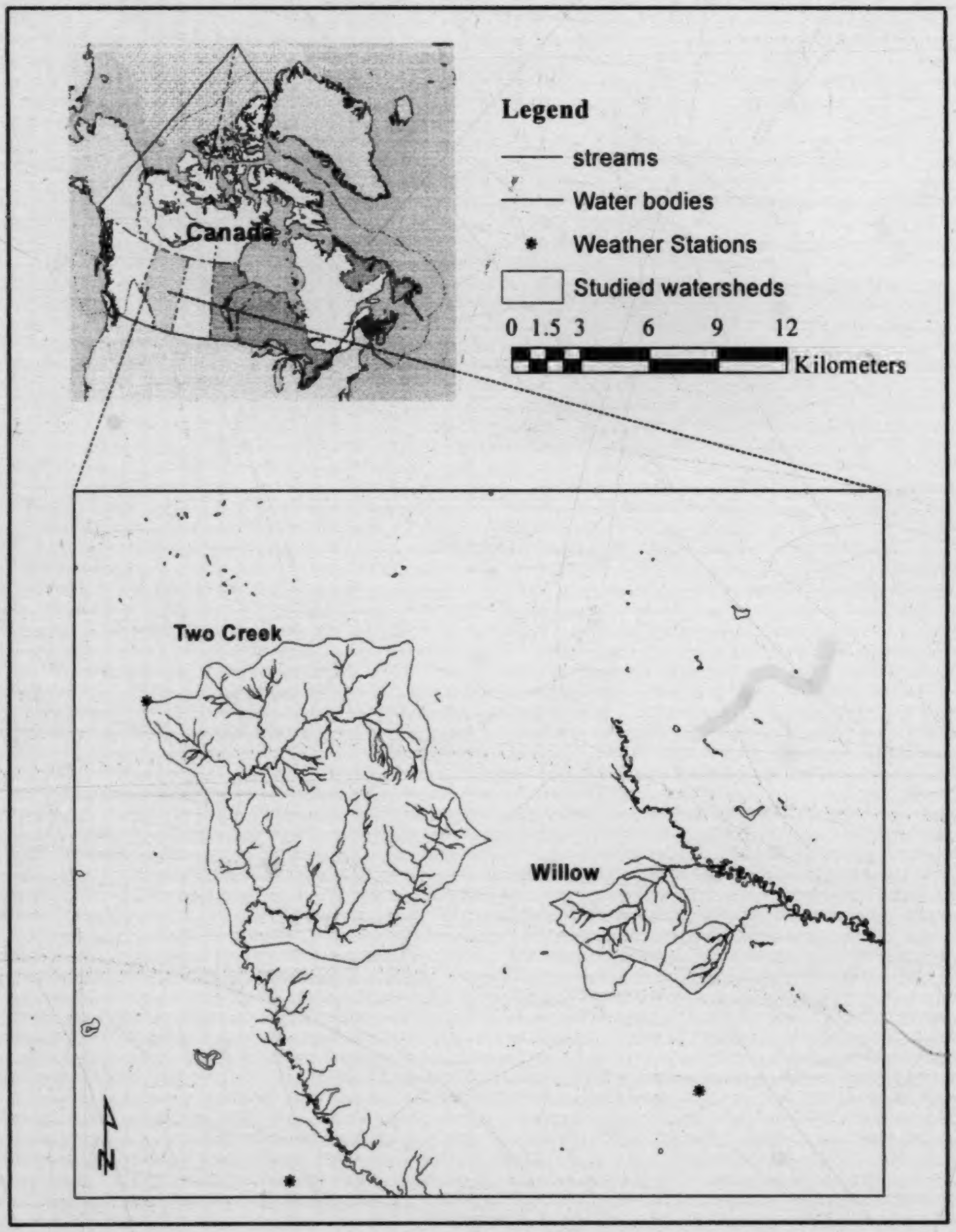

Canada
Legend
streams
Water bodies
Weather Stations
Studied watersheds
0 1.5 3 6 9 12
Kilometers
Two Creek
Willow
N

Figure 7-1. Study Area: Willow Creek and Two-Creek watersheds, Alberta, Canada

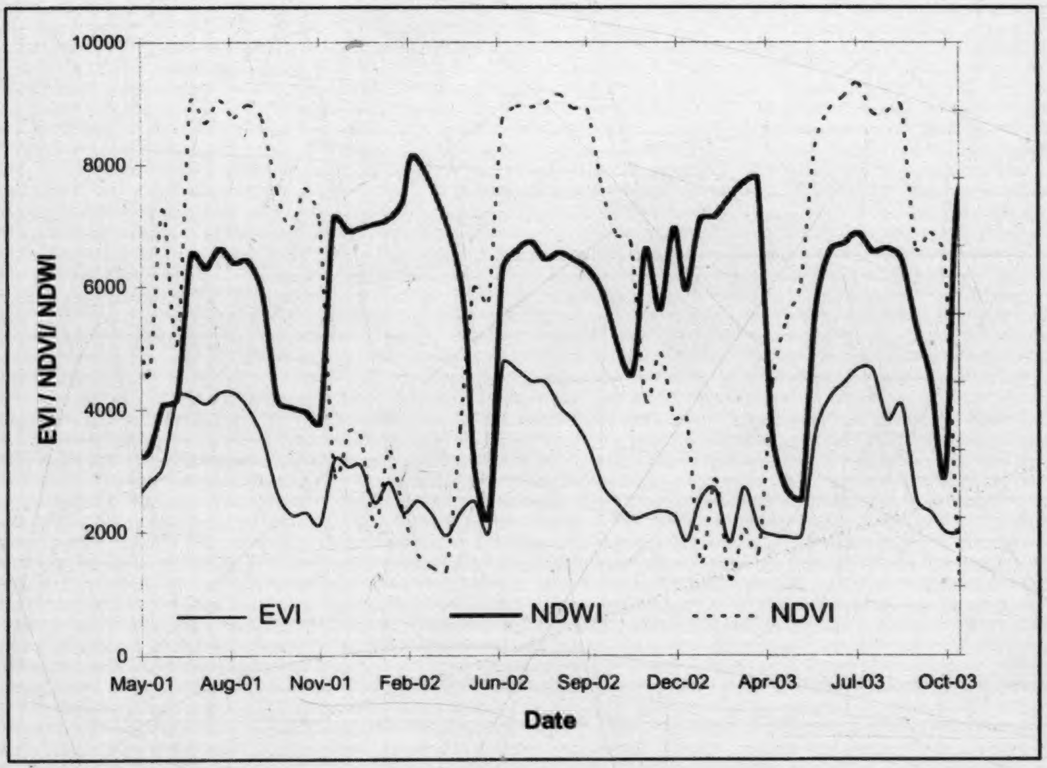


Figure 7-2. Comparison of EVI, NDVI, and NDWI time profiles for The Willow Creek watershed

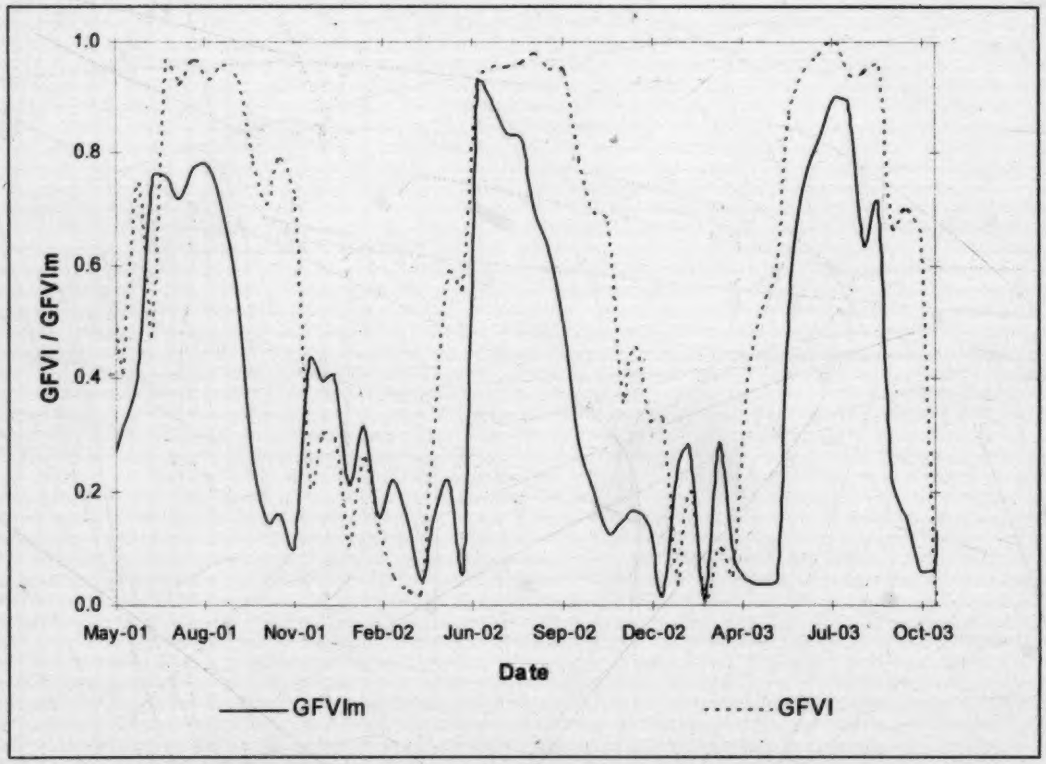


Figure 7-3. Comparison of GFVI and GFVIm time profiles for the Willow Creek watershed

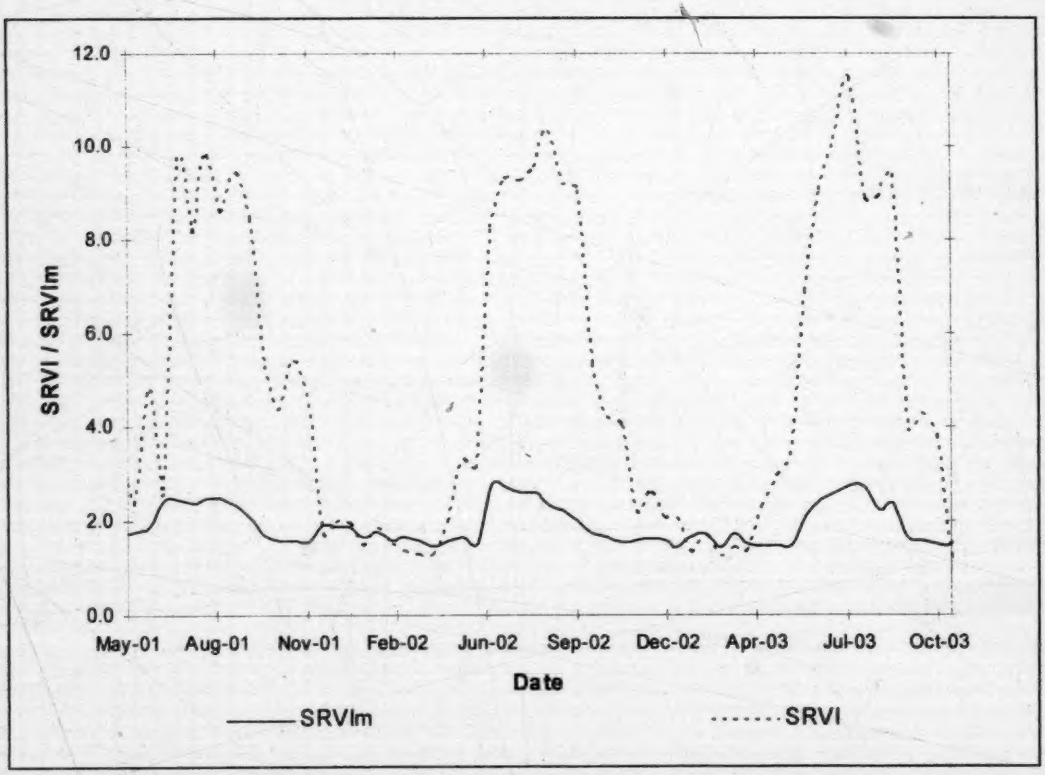


Figure 7-4 . Comparison of SRVI and SRVIm time profiles for the Willow Creek watershed

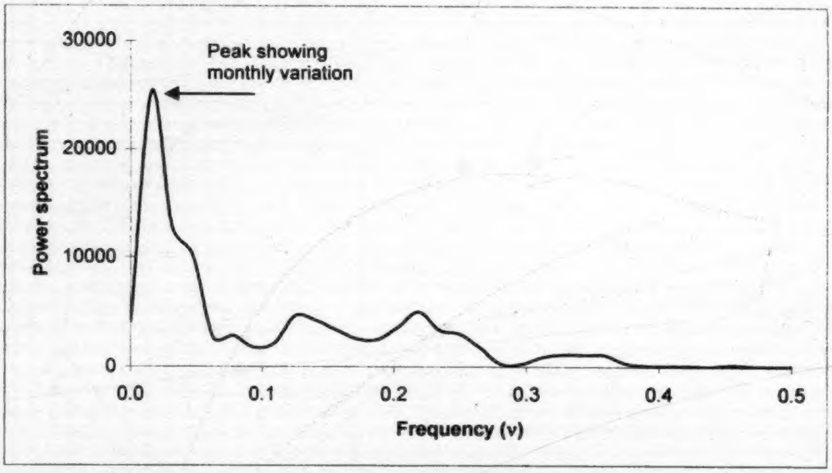


Figure 7-5. Power spectrum of TP in $[\mu g/l]^2$ for the Willow watershed

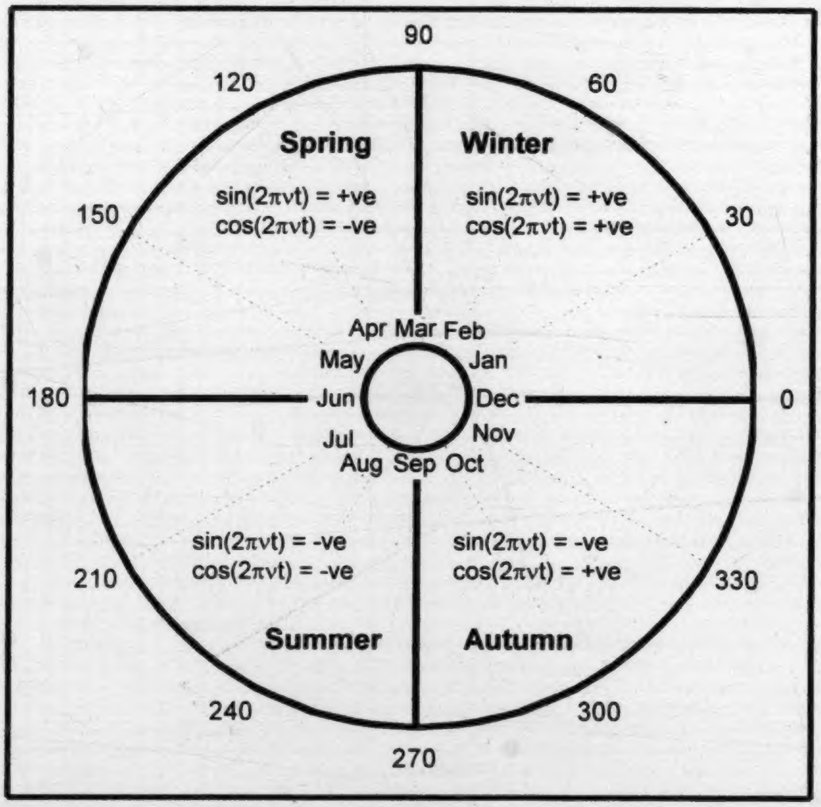

90
120
60
150
30
180
0
210
330
240
300
270
Spring
Winter
Summer
Autumn
sin(2πνt) = +ve
cos(2πνt) = -ve
sin(2πνt) = +ve
cos(2πνt) = +ve
sin(2πνt) = -ve
cos(2πνt) = -ve
sin(2πνt) = -ve
cos(2πνt) = +ve
Apr Mar Feb
May
Jan
Jun
Dec
Jul
Nov
Aug Sep Oct

Figure 7-6. Trigonometric quadrants describing the concept of feeding ANN TP models with seasonal variation and TP/Q hystereses loops

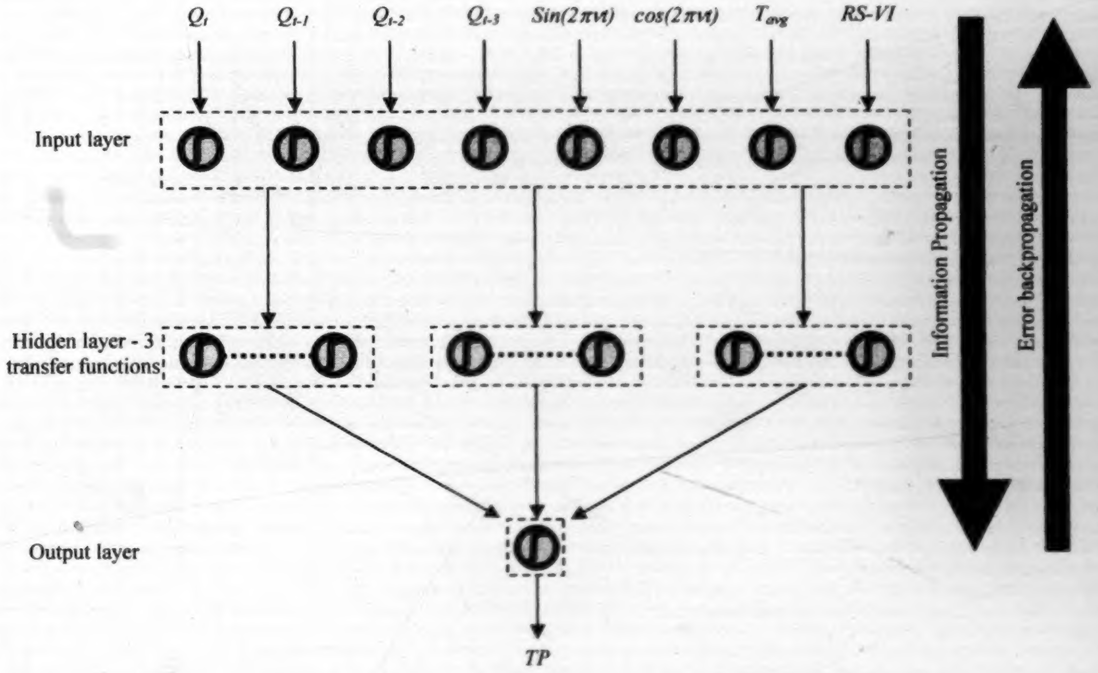

Q_t
Q_{t-1}
Q_{t-2}
Q_{t-3}
$Sin(2\pi vt)$
$cos(2\pi vt)$
T_{avg}
RS-VI
Input layer
Hidden layer - 3 transfer functions
Output layer
TP
Information Propagation
Error backpropagation

Figure 7-7. A schematic showing ANN optimum architecture for all six models

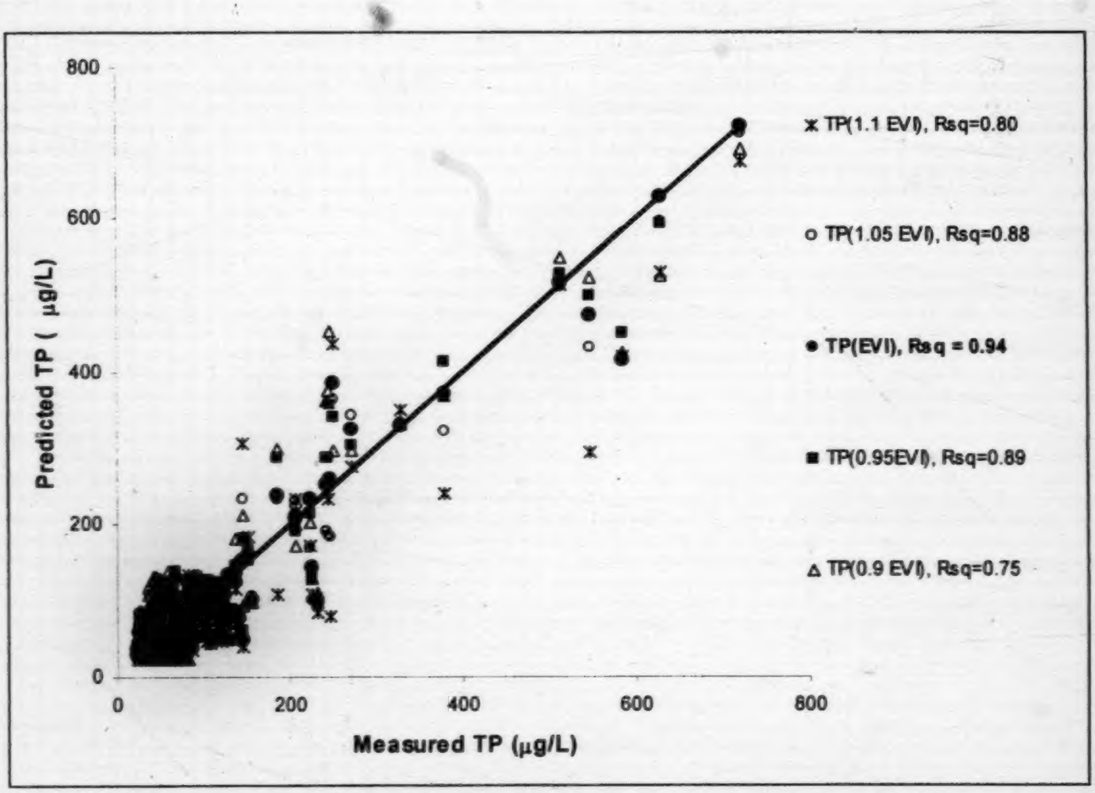


Figure 7-8. Sensitivity of TP predictions to changes in EVI, Rsq = R^2

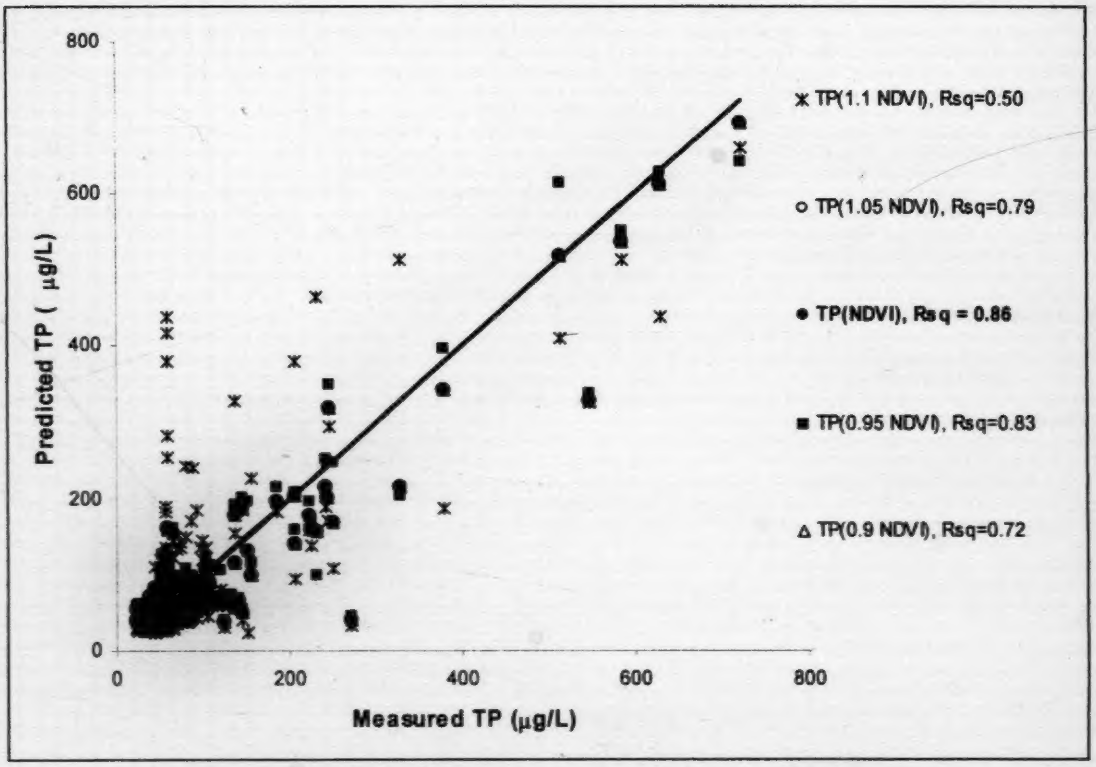


Figure 7-9. Sensitivity of TP predictions to changes in NDVI, Rsq = R^2

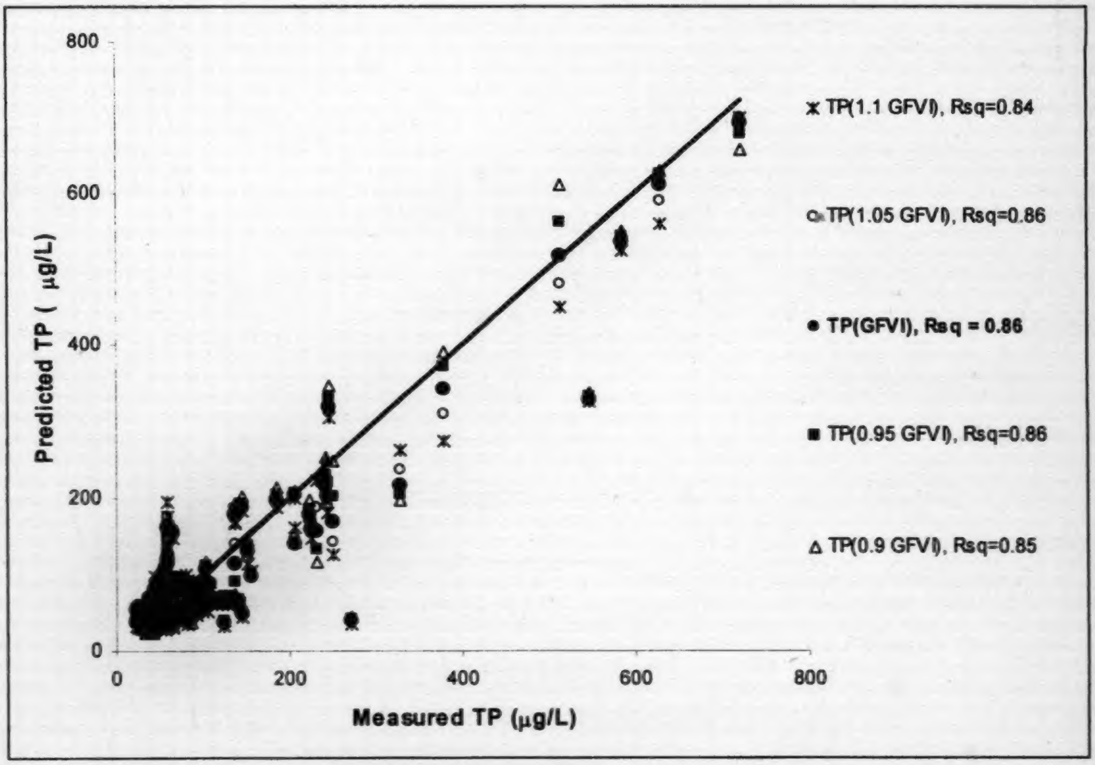


Figure 7-10. Sensitivity of TP predictions to changes in GFVI, Rsq = R^2

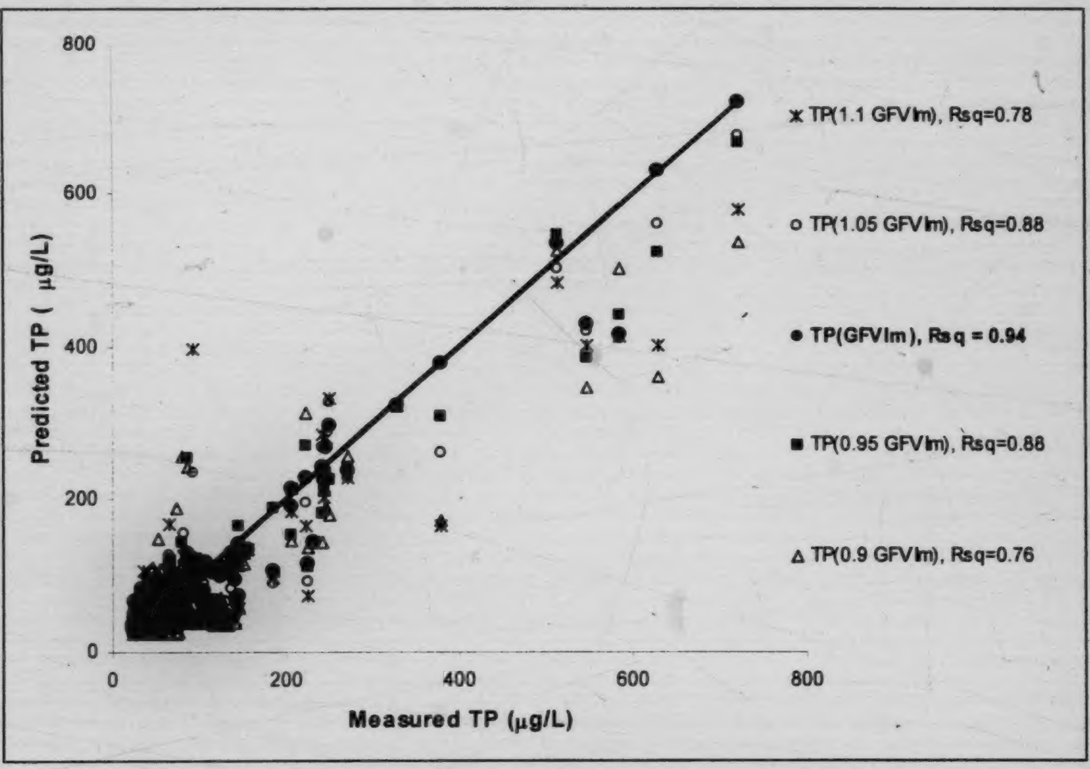


Figure 7-11. Sensitivity of TP predictions to changes in GFVIm, Rsq = R^2

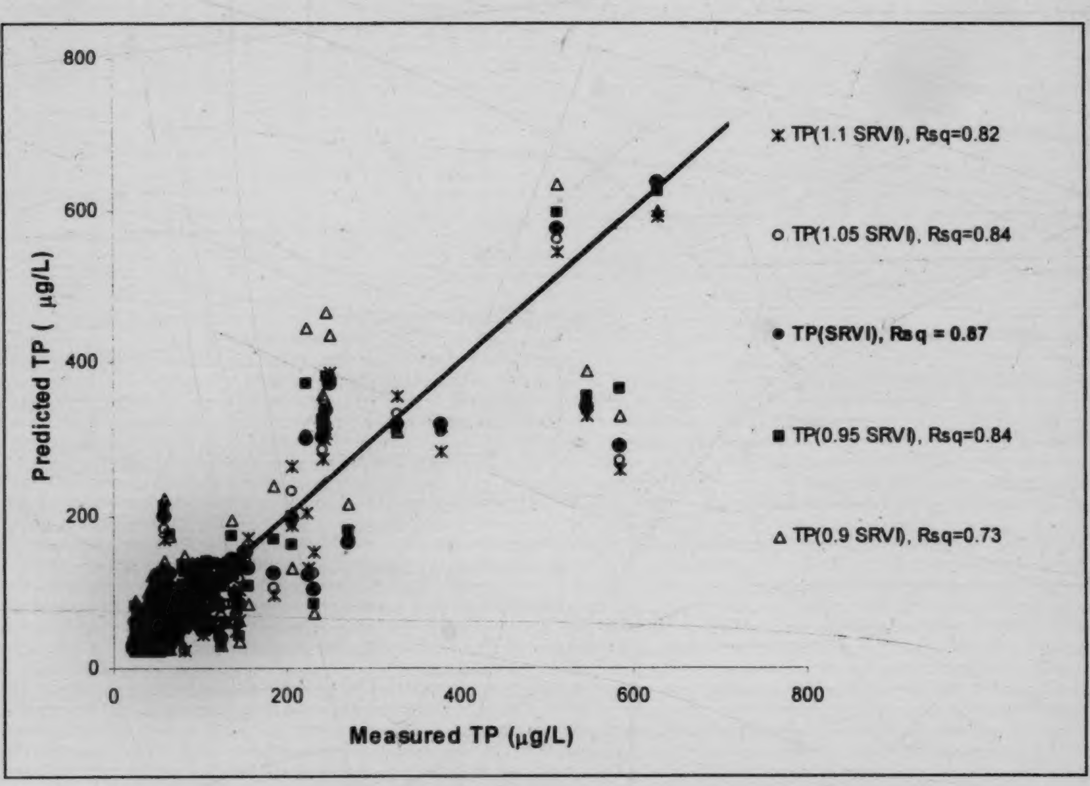


Figure 7-12. Sensitivity of TP predictions to changes in SRVI, Rsq = R^2

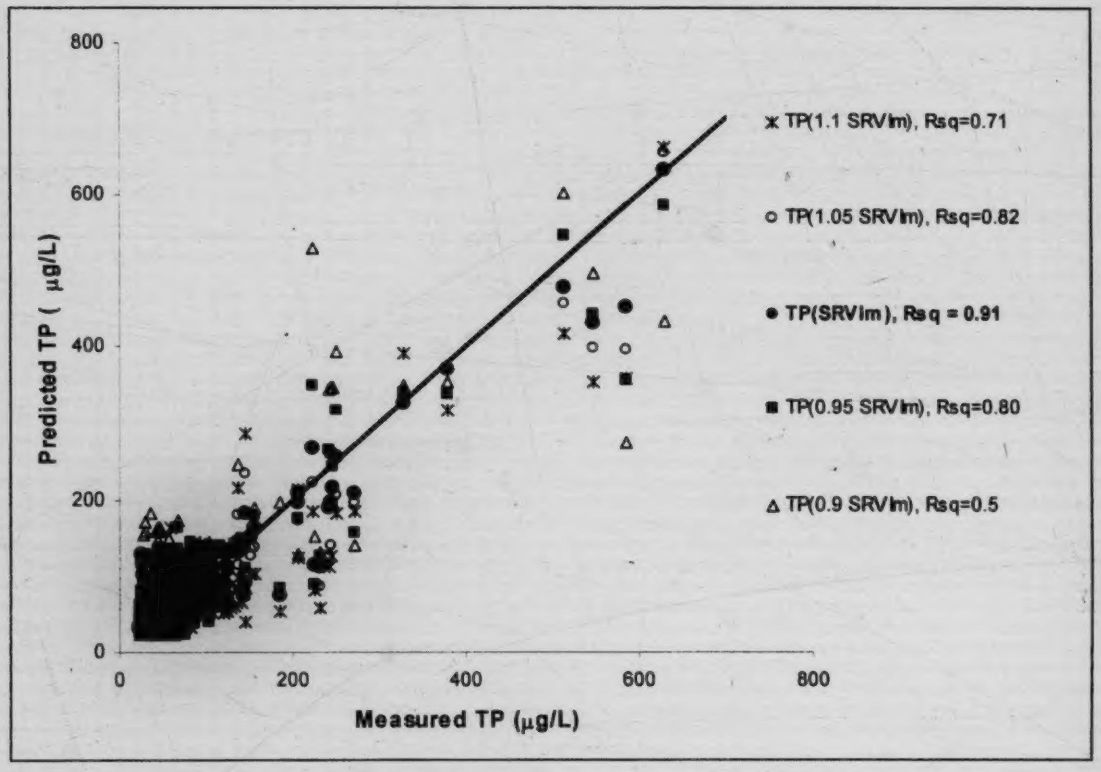


Figure 7-13. Sensitivity of TP predictions to changes in SRVIm, Rsq = R^2

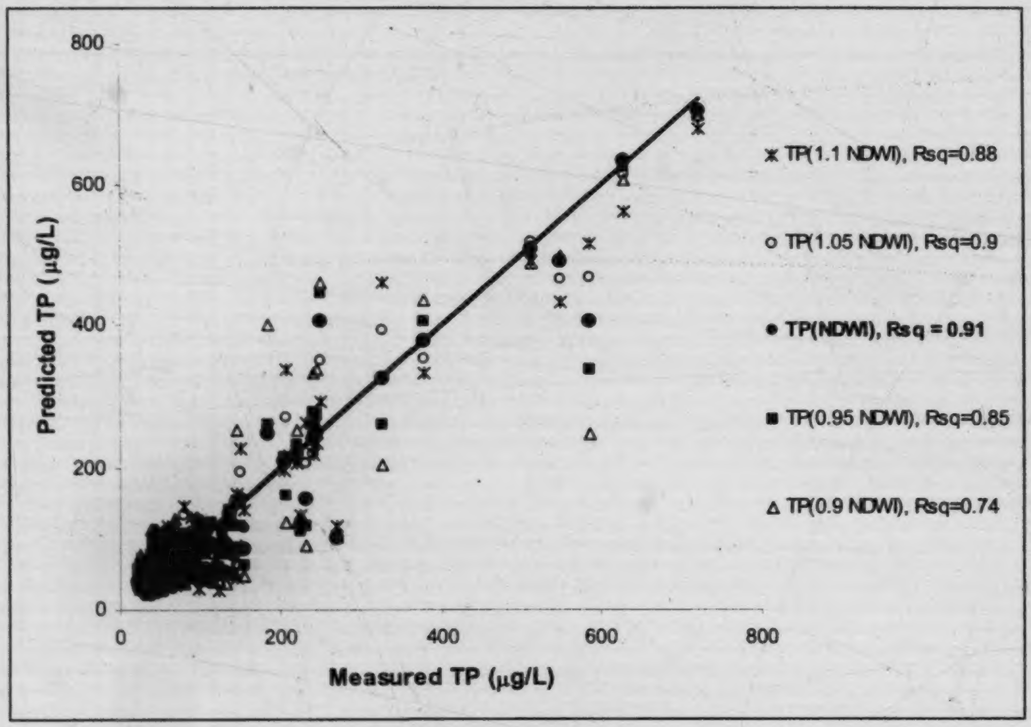


Figure 7-14. Sensitivity of TP predictions to changes in NDWI, Rsq = R^2

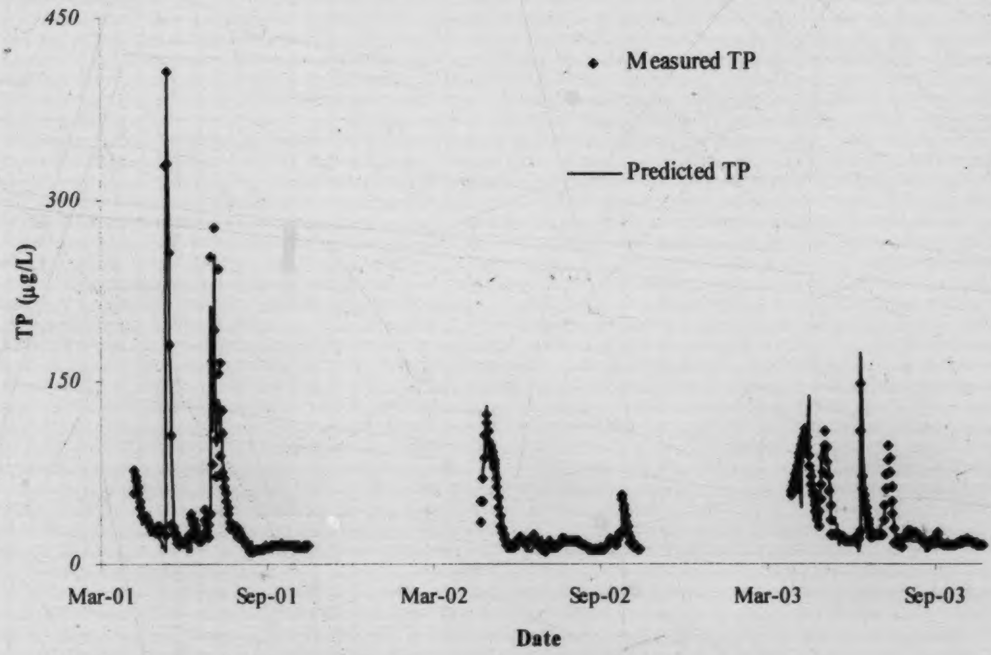


Figure 7-15. Measured vs. ANN predicted TP concentration profiles for the ANNTP(EVI) model applied to the Two-Creek watershed

CHAPTER 8. EFFECT OF WATERSHED SUBDIVISION ON WATER-PHASE PHOSPHORUS MODELLING: AN ARTIFICIAL NEURAL NETWORK MODELLING APPLICATION

8.1. Introduction

Natural and anthropogenic watershed disturbances can alter the hydrologic budgets and may expose soils to erosion, resulting in the potential for increased export of nutrients to surface waters (Smith et al. 2003; Munn and Prepas 1986; Cooke and Prepas 1998). The resulting excessive nutrient loads can cause an imbalance in biomass production in an aquatic ecosystem. The system then reacts by producing more phytoplankton than can be consumed by the ecosystem. This overproduction can lead to a variety of problems ranging from anoxic waters (through decomposition) to toxic algal blooms and a decrease in habitat diversity, thus leading to habitat destruction (Chorus 2001; Landsberg 2002; Hallengraeff 1993). Algal blooms' impacts can adversely affect not only the health of people, animals, and marine organisms, but also the "health" of local and regional economies (Hoagland et al. 2002). Hence, nutrient modelling, and in particular phosphorus (P) modelling, is critical to provide the necessary information for responsive watershed management practices.

A version of this chapter has been submitted for publication. Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. Effect of watershed subdivision on water-phase phosphorus modelling: an artificial neural network modelling application. *J. Environ.Eng. Sci.* (submitted 05/2007).

Numerous conceptual and data-driven models have been developed to simulate water-phase P dynamics. Conceptual watershed-scale P models include; but are not limited to; the soil and water assessment tool (SWAT) developed by Arnold et al. (1998), aerial non-point source watershed environment simulation-2000 (ANSWERS-2000) (Bouraoui and Dillaha 1996; Beasley et al. 1980), the hydrologic simulation program fortran (HSPF) (Johanson et al. 1984), the erosion productivity impact calculator (Sharpley and Williams 1990), the annualized agriculture non-point source pollutant loading model (AnnAGNPS) (Bingner et al. 2001), and the Guelph model for evaluating the effects of agricultural management systems on erosion and sedimentation (GAMES) by Cook et al. (1985). The use of this class of models presents the challenge of estimating or calibrating a large number of model parameters from limited available information. Obtaining necessary information for model calibration is always time consuming and expensive. Thus, most of the currently available models are limited in practice because of the extensive requirement for landscape data (e.g., data about soils, vegetation, and precipitation) needed for model calibration (Liu et al. 2006; Hauhs et al. 1996; Haan 1989).

Alternatively, data-driven models, in particular artificial neural network (ANN) models, have been successful in capturing data patterns without incorporating an extensive knowledge of the biological, geological, chemical, and physical behaviours of the system and, consequently, are attractive alternatives to traditional conceptual models when limited landscape data is available (Holmberg et al. 2006; Lek et al. 1996).

Watershed-scale P models can vary in complexity from lumped to spatially distributed representations of the landscape. Although the use of distributed models is conceptually appealing, the superiority of the more complex distributed models over the simpler lumped models is an issue of debate (Donnelly-Makowecki and Moore 1999; Wilcox et al. 1990; Michaud and Sorooshian 1994). Lumped models have minimal requirements for basin-specific data but cannot explicitly address the future impacts of changes in landscape activities. Distributed models, on the other hand, can explicitly account for spatial variability in the physical characteristics of a basin and, in principle, should perform better than lumped models when applied to ungauged basins or for predicting the impacts of land use changes (Refsgaard and Knudsen 1996). In practice, however, the superiority of more complex models over simpler ones for operational purposes is still in doubt (Donnelly-Makowecki and Moore 1999; Hauhs et al. 1996).

Several studies have addressed the impact of watershed subdivision on streamflow simulation. Most of these studies demonstrated that beyond an optimum threshold in terms of the number of subwatersheds, no significant improvement occurs in the modelling accuracy as a result of increasing the number of subwatersheds (Norris and Haan 1993; Zhang and Montgomery 1994; Kalin et al. 2003; Boyle et al. 2001). Other researchers have found that runoff volume is not significantly affected by the number and the size of subwatersheds (Tripathi et al 2006; Jha et al. 2004; Bingner et al. 1997; FitzHugh and McKay 2000). In contrast, Mamillapalli (1998) and Zhang et al. (2004) found that model runoff simulations tended to be more accurate with finer discretization of the watershed into subwatersheds. Thus, the role of spatial discretization on streamflow prediction is still unclear, with conflicting viewpoints being expressed by researchers. In addition, none of the above efforts considered the impact of watershed discretization on total phosphorus (TP) prediction accuracy.

This paper addresses the impact of watershed subdivision in modelling TP concentration. The understanding of this impact is important for judging whether the increased costs and effort of obtaining and processing spatially-distributed basin information can be justified in terms of the increased accuracy and reliability of model predictions.

In the current study, models were formulated with artificial neural networks (ANNs). To alleviate the typical problem of landscape data scarcity, our models were formulated using public domain remote sensing (RS) information and regularly available weather station data.

Satellite RS has recently made available a large inventory of cost-effective landscape data over the entire landbase, rather than providing only a sampling of it as is the case with ground-based measurements. The Moderate-resolution Imaging Spectroradiometer (MODIS) launched by the National Aeronautics and Space Administration (NASA) in December, 1999 has provided scientists with the ability to measure forest growth with greatly improved spatial and temporal resolution. Furthermore, MODIS data is freely available, thus providing a means of acquiring time series representation of vegetation dynamics at an affordable cost. The present study involves an attempt to build models that can rely on a dynamic suite of remotely sensed vegetation indices for predicting TP concentration in a 2nd order watershed. To our knowledge, it is the first to attempt to address the impact of watershed discretization on TP prediction. Earlier studies that utilized ANN for TP modelling focused on the lumped parameter representation of the modelled basin (Holmberg et al. 2006; Nour et al. 2006a; Lek et al. 1996). Thus, this study is the first to address the impact of watershed subdivision on a water quality parameter by using an ANN modelling algorithm.

The specific objectives of this work were: (1) to model TP as an influencing parameter for aquatic ecosystems on the Boreal Plain, (2) to develop a useful modelling tool that is less reliant on ground based data, (3) to utilize RS data availability via MODIS to represent vegetation dynamics in model formulation, and (4) to assess the impact of watershed subdivision on model performance.

8.2. Study Area and Input Database

The Willow watershed covers an area of 15.6 km^2 in the Virginia Hills of Northern Alberta, Canada (Figure 8-1). The area is exemplified mainly by low topographic relief and alkaline phosphorus-rich soils (mainly fine-textured Luvisols) developed from sedimentary bedrock. The studied watershed is mostly forested. Its forests contain white spruce (Picea glauca), lodgepole pine (Pinus contorta), trembling aspen (Populus tremuloides), and balsam poplar (P. balsamifera). The climate is cool-temperate (the mean monthly air temperatures range from –23 to 18 °C), and the mean annual precipitation is 584 mm (1972 to 1997 (Environment Canada 2002)). The study watershed is one of the control watersheds of the Forest Watershed and Riparian Disturbance (FORWARD) Study (see *J. Environ. Eng. Sci.* special issue Volume 2, 2003 for details). The data used for formulating, calibrating and validating the study watershed can be divided into two main sources: ground based and remotely sensed.

8.2.1. Ground-Based Data Acquisition

Average daily air temperature data was obtained from the Environment Canada weather station located at the Whitecourt airport (Figure 8-1). Daily stream flow (Q) and TP concentration data were collected from May 2001 to October 2004 as part of the FORWARD Study. A 30 m x 30 m resolution digital elevation model (DEM) was acquired for the study area (Figure 8-2) and was used for watershed delineation.

8.2.2. Remotely-Sensed Data Acquisition

Watershed level vegetation dynamics were captured by using the Moderate Resolution Imaging Spectroradiometer (MODIS) atmospherically corrected reflectance data. This information was further manipulated to produce remotely sensed vegetation indices (VIs). Two vegetation indices were used in the TP modelling as surrogates for the soil/vegetation phosphorus transport. The first was the enhanced vegetation index: EVI (Eq. 1), a remote sensing (RS) representation of vegetation chlorophyll content that was designed to perform better than its predecessors in dense vegetation (Huete et al. 2002). The second index was the normalized difference water index (NDWI) represented by Eq. 2., which is a RS representation of vegetation water content (Cheng et al. 2006).

$$[1] \quad \mathrm{EVI} = G\frac{\lambda_{NIR} - \lambda_{red}}{\lambda_{NIR} + C_1 \times \lambda_{red} - C_2 \times \lambda_{blue} + L}$$

$$[2] \quad \mathrm{NDWI} = \frac{\lambda_{NIR} - \lambda_{MIR}}{\lambda_{NIR} + \lambda_{MIR}},$$

where, λ_{red}, λ_{blue}, λ_{NIR}, and λ_{MIR} are the atmospherically corrected surface reflectance at the red, blue, near infra red, and mid infra red wavelengths; L is the canopy background adjustment (to correct for non-linearity, differential NIR and red radiant transfer through a canopy); C_1 and C_2 are coefficients of the aerosol resistance term, which use the blue band to correct for aerosol influences in the red band; and G is the gain factor. The coefficients adopted in the EVI algorithm are $L = 1$, $C_1 = 6$, $C_2 = 7.5$, and $G = 2.5$ (values after Huete et al. 2002).

MODIS has the highest spectral resolution compared to other global coverage moderate resolution spectroradiometers and provides better cloud and atmospheric characterization (Justice et al. 1998). While viewing the entire earth's surface every 1 to 2 days and acquiring data in 36 spectral bands, it has an improved spatial resolution of 250 m. The dataset comes in the Hierarchical Data Format - Earth Observing System (HDF-EOS), which is the standard archive format for EOS Data Information System (EOSDIS) products. HDF-EOS is a multi-object file format and supports a variety of data types. Each HDF-EOS file size is approximately 500 MB and consists of 11 Science data sets (SDSs), which are the actual data stored in an array format (MODIS Land Science Team (MLST) 2004)). The first two layers of an HDF-EOS file are the EVI and NDVI images. These are 16-day composite, re-sampled, 250 m spatial resolution, cloud-free, pre-processed high-quality imagery pixels produced for each year since 2000. Although the valid range of NDVI or EVI is from -0.2 to +1, the values are scaled up by a factor of 10,000 with a fill value of -3000. The next two layers provided the NDVI and EVI per-pixel quality information followed by the four SDSs of reflectance data (red (620 to 670 nm), NIR (841 to 876 nm), blue (459 to 479 nm), and

MIR (2105 to 2155 nm)) from which the NDWI and EVI were derived. The last three SDSs provide the sun-canopy-sensor angles. This information is needed to know the variable scan geometry under which the pixel reflectances were measured by the MODIS sensor. Each SDS (or layer) is a tile unit (fixed-area size) in the Sinusoidal (SIN) grid projection. The tile unit is the smallest unit of MODIS land data processed at any time and has an aerial extent of approximately 1200 km x 1200 km.

A sequence of four years, from 2001 to 2004, of the MODIS HDF-EOS files was ordered through the EOS data gateway interface. The gateway can be accessed by using MODIS' website: http://modis.gsfc.nasa.gov/ through the "data" link. The Terra MODIS MOD13Q1 dataset was chosen in the search criteria, and the coordinates of the study area were fed into the system. The search criteria retrieved the datasets from the years 2001 to 2004. Twenty-three EOS-HDF files for each year (a total of 92 files with 92 metadata files) were downloaded from NASA's server. After quality checks, 88 MOD13Q1 files were imported into ERDAS Imagine® GIS software. All 11 Scientific Data Sets (SDSs) produced by MODIS in each HDF-EOS file were converted to 11 image files (.img), the ERDAS Imagine® standard raster file format. For each band of interest (red, blue, NIR, MIR, and EVI), the images for each year were then stacked by using the "Image Stack" module and subsetted by using the Area of Interest (AOI) shape file of the studied watershed, or its subwatersheds according to the case. Finally, each SDS was averaged over the watershed area, or its subwatersheds according to the case, to present the selected area's overall response with respect to vegetation dynamics. This information was then exported to a database. Equations 1 and 2 were then applied to the respective reflectance information to evaluate the EVI and NDWI, respectively.

The results were then interpolated by using spline interpolation to obtain a value at each time instant when the TP concentration was measured in the stream. This information was finally used as inputs to the devised ANN models.

8.3. Watershed Delineation/Discretization

Automatic delineation of the watershed was achieved using a 30 m x 30 m resolution raster DEM of the FORWARD study area. The commonly used D8 method (Fairchild and Leymarie 1991) was used to determine the direction of the overland flow at each cell of the DEM. ArcHydro extension of the ArcGIS 9.1 package from ESRI was used for watershed delineation and further subdivision into smaller subwatersheds. First, the input DEM was modified to remove depressions and flat areas, thus eliminating indefinite down-slope drainages (Figure 8-2). Second, by using the D8 method, the flow direction in each grid cell was estimated by allowing the water to flow from each cell to only one of the eight neighbouring cells, along the direction of the steepest descent. Each grid cell was then given a value corresponding to its flow direction. The output of this operation was a raster image with eight distinct values according to the flow direction, as represented by Figure 8-3. Third, the number of cells draining into a given cell along the flow network was monitored and reported at the drained-to cell formulating the flow accumulation raster map (Figure 8-4). The flow accumulation grid can thus be viewed as the drainage area reported in units of numbers of grid cells. As expected, the large values of "flow accumulation" are associated with streams receiving water from the surrounding land. Fourth, streams were defined by identifying a threshold drainage area (TDA). All cells with a "flow accumulation" value greater than or equal to the specified TDA were classified as stream cells, whereas the others were considered as land cells draining into the stream cells. The stream cells were assigned a value of "1", while the land cells were assigned a value of "NO DATA." Fifth, stream links were generated by separating each stream at the confluence point and giving each stream a distinct value. The zone cells, whose drainage flows into each stream link, were identified and given the value of the stream link they drain into. Each delineated catchement grid was then transferred to a polygon by using the ArcGIS "raster-to-vector" tool formulating the delineated basin. Figure 8-5 depicts the delineated catchments for the FORWARD study area for a representative TDA of 250 ha. Sixth, several TDAs were used to generate the corresponding catchments in order to identify the impact of watershed subdivision on the modelled TP concentration. The choice of a good TDA is rather arbitrary and case-sensitive. However, it is common practice to start with a big TDA—that will yield a single representation of the watershed—and then to reduce the TDA until spurious (small and/or highly elongated) subwatersheds start to appear (FitzHugh and Mackay 2000). This approach was implemented in the current study to produce four watershed subdivisions, ranging from no subwatersheds (i.e., lumped parameters representation) to 11 subwatersheds, as shown in Figure 8-6.

8.4. Methodology

The previous efforts to model water-phase TP concentrations showed that there is a need for information regarding soil/vegetation phosphorus content to reasonably model TP concentration. Without this information, only the daily change in the TP concentration can be accurately predicted (Nour et al. 2006a). This study attempted to devise a robust ANN TP model that relies on remotely sensed vegetation information to represent the missing soil/vegetation phosphorus component in the previous model described by Nour et al. (2006a). The goal was to build models that could rely on available information in the Boreal forest of Canada. Building models that do not rely on watershed-specific measurements would help to make the developed models applicable to similar watersheds without the need to invest in collecting watershed-specific data. The devised models relied solely on weather station information available from Environment Canada and local fire towers, in addition to public domain remote sensing information. The impact of watershed subdivision on the prediction performance of the developed models was studied in detail. The following steps summarize the methodology used to achieve the study objectives:

- Modelling TP in ungauged watersheds requires utilizing modelled flow values —in place of measured ones—as inputs. Thus, the flow model proposed by Nour et al. (2007) was used to generate a time series of daily streamflow values for the studied watershed. This flow model relies only on available weather station information and, therefore, can be used with the currently available input database in the Boreal forest.
- Four watershed delineations were achieved to test the impact of watershed subdivision on TP predictions. For each delineation, a time series of MODIS derived EVI and NDWI was calculated for each subwatershed.
- Four ANN TP models were developed. All models utilized the same inputs except for the inputs representing the VIs. For these inputs, a semi-distributed

representation of the EVI and NDWI was allowed to test the impact of watershed subdivision on model performance.

- The relative contributions of model inputs to TP predictions was computed for the four models and compared to assess model consistency. Finally, the spatial distribution of the relative contribution of the landscape vegetation content to the water-phase TP was studied to help guide forest management in the region.

8.5. ANN Model Building

ANN applications have expanded over the past two decades to include such disciplines as engineering, computer science, statistics, physics, medicine, biology, pharmacy, and psychology. In these fields, ANNs are beginning to be the favoured option over other modelling alternatives because ANNs are highly nonlinear and universal approximators. The main advantages of ANNs are their ability to model nonlinear processes of a system without any *a priori* assumptions about the nature of the generating processes, and their ability to efficiently handle incomplete noisy and nonstationary data (Zealand et al. 1999).

The multi-layer perceptron neural network trained with the error back-propagation training algorithm (MLP-BP) is, by far, the most popular of all neural networks (Dawson and Wilby 2001; Maier and Dandy 2000). Because of the popularity of the algorithm and because the authors have applied it successfully in different applications, it was used in this study.

To construct a robust ANN model, three stages were implemented in this study: (1) data pre-processing, (2) model formulation, and (3) model assessment. Each stage is briefly described in the following sections.

8.5.1. *Data Pre-processing*

In the data pre-processing phase, the input variables were identified; all input/output data patterns were explored to highlight the data's main trends and features; and any irregularities entrenched in the data were checked further to ensure good data quality. Data pre-processing involved determining the appropriate model inputs, and dividing the data patterns into calibration, validation, and cross-validation data sets.

8.5.1.1 Determining appropriate model inputs

The main objective of this piece of work was to build TP ANN models that could rely on readily available information on the Boreal plain of western Canada. Thus, the input selection was based on our conceptual understanding of the process under investigation, but was constrained by data availability. Nour et al. (2006b) constructed an ANN model for modelling the daily change in TP concentration and concluded that supplementary information regarding soil/vegetation phosphorus content is needed to facilitate the modelling of daily TP concentrations. They recommended dividing the inputs into causal inputs, time-lagged inputs, and inputs reflecting TP/Q hystereses loops. A similar approach was used here by adding a fourth component to represent the subdivision of the watershed under investigation.

The causal inputs included streamflow (Q), average air temperature ($T_{avg.}$), and both the NDWI and EVI as surrogates for the soil/vegetation phosphorus content.

To account for the time-series effect of the inputs, time-lagged inputs were incorporated in the model by applying a cross-correlation analysis as described in Nour et al. (2006b) to identify possible time-lagged inputs. This step resulted in including Q three days ahead (Q_{t-3}), Q two days ahead (Q_{t-2}), and Q one day in advance (Q_{t-1}) in addition to Q_t as inputs representing the daily streamflow.

To take into consideration the inputs reflecting TP/Q hystereses, spectral analysis, as proposed by Nour et al. (2006b), was conducted to identify the dominant frequency identifying the cyclic/seasonal fluctuations. A smoothed periodogram was constructed, and a monthly variation was found to be the main contributor to the cyclic fluctuations. The TP/Q hystereses was then incorporated in the model by adding two additional inputs, $\sin(2\pi \frac{t}{12})$ and $\cos(2\pi \frac{t}{12})$. By looking at the sign of the two inputs, the model is expected to be able to identify the season under study (for example, a negative value for both the sine and the cosine inputs identifies the summer season), and with the aid of the magnitude of these two inputs, the model is believed to be able to identify the month within each season, as illustrated by Figure 8-7.

Since semi-distributed models had to be formulated, the EVI and NDWI were evaluated for each subwatershed and were included as multiple inputs corresponding to the number of subwatersheds under study (for example in the TDA = 250 ha model, five inputs were used to represent the EVI of the five subwatersheds, and five inputs were used for the NDWI variables of the five subwatersheds).

The previous categorization of inputs resulted in using 9, 17, 21, and 29 inputs for the lumped, TDA = 250 ha, TDA = 100ha, and TDA = 50 ha models. Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, $\sin(2\pi\frac{t}{12})$, $\cos(2\pi\frac{t}{12})$, and $T_{avg.}$ were common inputs to all four models. The number of EVI and NDWI inputs varied and was equal to twice the number of subwatersheds in each model.

8.5.1.2 Data division

Three data sets were used for a rigorous analysis of a candidate ANN model. The first set is the training data set (TS), which was used for model training and the optimization of the connection weights. The second set is the testing set (SS), which was used to decide when to stop training to avoid model overfitting. Thirdly, the cross-validation data set (CVS) was used to evaluate the model against a totally independent data set. In this study, the data patterns were divided into three data sets in the ratio of 3:1:1 for training, testing, and cross-validating the model, respectively. The split was based on an algorithm that targeted a similar frequency distribution of each data set, with any extreme and rare values being assigned to the training data set.

8.5.2. Model Formulation

The methodical approach suggested by Maier and Dandy (2000) and Gamal El-Din and Smith (2002) was adopted in this study with some modifications. Earlier work found that a typical feed-forward (FF) multi-layer perceptron (MLP) ANN with a single hidden layer utilizing one activation function in its processing elements (nodes) cannot accurately map the streamflow at higher latitudes (Nour at al 2006b). Thus, a modification was made by applying a FF-MLP ANN with one hidden layer with processing elements that used more than one activation function. The approach proposed in Maier and Dandy (2000) and Gamal El-Din and Smith (2002) was then used to optimize the number of nodes and their corresponding activation functions.

Two training algorithms were tested in this study: (1) the typical gradient descent BP algorithm that uses a learning rate and a momentum coefficient to control the training speed and to facilitate moving towards a global minimum in the error surface (Haykin 1994), and (2) a BP algorithm with a batch update (BP-BU) technique. The NeuroShell 2 software package from Ward Systems Group was used to train the models. In BP learning, minimizing an error function, the mean squared error (MSE) in this case, is desirable; however, this process is complicated due to a typically multi-local minima error surface. Attempting to build a robust model that would not become overfamiliarized with the training data, to the extent that it could not generalize to problems it had not yet encountered, we used the early stopping technique that relies on the testing data set statistical performance measured by the MSE to dictate when to stop training. Typically, the repeated training iterations successively enhance the network's performance in the training data set, but the testing data set performance has an optimal point beyond which the statistical performance deteriorates again. Training continued as long as the error of the testing data set was continuously decreasing, and was halted when this error started to increase, even if that of the training data set were still decreasing, but the connection weights were always adjusted based on the training data set.

Applying the abovementioned model development algorithm to each of the four developed models yielded similar model architectures with the MLP-BP using one hidden layer with three activation functions (namely Gaussian, Gaussian complement, and logistic functions), a linear scaling function, and a hyperbolic tangent output activation function. The typical gradient descent BP algorithm that uses a learning rate and a momentum coefficient was found to perform better than the BP-BU method. Table 8-1 summarizes the optimum ANN architecture including the network internal parameters for the four devised models. The similarity in the architecture for all models highlights the strength and consistency of our modelling algorithm. Figure 8-8 depicts a general ANN architecture that can represent any of the devised models.

8.5.3. Model Assessment

There is a general agreement in the literature that one should not rely on an individual error measure when assessing ANN model performance (Dawson and Wilby 2001; Legates and McCabe 1999). Thus, a number of complementary error measures were used in this study: (1) percent relative bias, RB(%) (Eq. 3); (2) the root mean squared error, RMSE (Eq. 4); (3) the mean absolute error, MAE (Eq. 5); (4) the square of Pearson's correlation coefficient, r^2; (5) the coefficient of multiple determination, R^2 (sometimes referred to as the coefficient of efficiency) (Nash and Sutcliffe 1970) as denoted by Eq. 6; (6) the second-order index of agreement, d_2 (Willmott et al. 1985) represented by Eq. 7; (7) the multivariate corrected Akaike's information criterion, AIC_c (Eq. 8) suggested by Bedrick and Tsai (1994); and (8) the Bayesian information criterion, BIC (Eq. 9) recommended by Schwartz (1978):

$$[3] \quad RB(\%) = 100\left(\frac{\sum_{i=1}^{N} P_i - \sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} O_i}\right)$$

$$[4] \quad RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2}$$

$$[5] \quad MAE = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - P_i\right|$$

$$[6] \quad R^2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$$

$$[7] \quad d_2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|)^2},$$

where P_i and O_i are the predicted and the measured TP concentration values at time *i*, respectively; N is the number of data points; and $\overline{O}$ is the mean of the measured TP concentration for the entire time period.

$$[8] \, AIC_c = \ln\left[\frac{RSS_k}{N}\right] + \frac{N+k}{N-k-2}$$

$$[9] \, BIC = \ln\left[\frac{RSS_k}{N}\right] + \frac{k\log(N)}{N},$$

where RSS_k denotes the residual sum of squares under the model with k parameters, and N is the number of data patterns.

The first six measures of "goodness-of-fit" were calculated for the training, the testing, and the cross-validation data sets and assessed to judge the candidate model's performance. One would expect that a model with more parameters would match data better than a model with fewer parameters; however, increasing model complexity does not necessarily lead to proportionate increases in model accuracy. Therefore, the last two performance measures were used because they penalize the models with more parameters and can thus provide a good evaluation of model parsimony when models are to be compared.

A graphical representation of the measured and the predicted TP concentration profiles was then investigated to highlight zones of poor performance. In addition, the possible reasons for the poor simulations of some data regions were identified, and model improvements were attempted.

8.6. Results and Discussion

Four ANN models were devised and applied to the Willow watershed to simulate daily TP concentrations over the period from May 2001 to October 2004. Eight statistical measures of model performance were used to assess the prediction accuracy of each model. The period of the study was divided, in each model, into three data sets for model calibration, testing, and cross-validation. The same data sets were used in all four models to provide a similar basis of comparison. The statistical measures of the "goodness-of-fit" of all candidate models are summarized in Table 8-2. All models produced an almost perfect mass balance of TP with a relative bias not exceeding 3% for all data cases. RMSE and MAE were reasonably low for all candidate models and all data sets. Both RMSE and MAE were decreased by increasing the degree of watershed subdivision. However, the difference in RMSE and MAE between the coarsest resolution and the finest resolution ranged from 4 to 10 μg/L for all training, testing, and cross-validation data sets, suggesting that all four models are reasonably accurate. The index of agreement, d_2, and the coefficient of multiple determination, R^2, exceeded 0.95 and 0.82, respectively, for all the studied data sets for the four developed models. The comparatively high values for all the training, the testing, and the cross-validation data sets, for all candidate models, reflects the superiority of the modelling approach and its good generalization ability. Both d_2 and R^2 increased in value by increasing the number of subwatersheds. However, all models provided satisfactory results in terms of "goodness-of-fit" statistics.

To better compare the candidate models, two measures of model performance (AIC_c and BIC), which have a penalty term for increasing model complexity were used to assess whether the increased prediction accuracy could justify the increased model complexity. The full period of the study was used as a basis for comparing the four models. The results showed that both AICc and BIC were decreased by increasing the number of subwatersheds, suggesting that Model 4 (TDA = 50 ha) is the most parsimonious (Table 8-2). Figure 8-9 depicts the changes in AICc, BIC, and R^2 with increasing the number of watershed subdivisions. It reflects the previously described trend of increasing R^2 and decreasing both AICc and BIC by incorporating more subwatersheds into the modelling exercise. However, the slope of enhancement decreased significantly after 7 subwatersheds, suggesting that the enhancement of the model performance between 7 subwatersheds and 11 subwatersheds was operationally minimal, yet statistically significant.

Although model 4 (TDA = 50 ha) was proved to be statistically the most parsimonious, it is important to visually compare the TP concentration profiles of all models attempting to compare the models' performance in all date ranges. The cross-validation data set was used as a basis of comparison to test the accuracy of all four models when applied to an independent data set. All models replicated the TP profile reasonably well (Figure 8-10). All models did not experience any lag phenomena (peak location shift) in prediction; however, the semi-distributed models managed to predict peak responses a little better than the lumped model (Figure 8-10).

In Alberta, the surface water quality guidelines for the protection of freshwater and aquatic life dictate a maximum allowable TP concentration of 50 μg/L (Alberta Environment 1999). However, Alberta forest management standards require maintaining such low levels of TP concentration only for third-order watercourses and higher (Alberta Sustainable Resource Development 2006). Although the studied watershed is lower than third-order, and thus is allowed to experience TP concentrations above 50 μg/L, it is important to assess the comparable models' performance in predicting TP concentration values above 50 μg/L, being of more importance to forest management. Thus, all data patterns of the cross-validation data set that had a TP concentration above 50 μg/L were clipped, and the comparative model accuracy in terms of the statistical measure of "goodness-of-fit" was computed. The results showed that the statistical performance measures included in Table 8-2 did not change considerably by focusing only on TP concentrations above 50 μg/L. To test the models' stability, the testing and the cross-validation data sets were swapped, and the models were retrained. Next, the performance measures were recalculated for the new cross-validation data set. All measures of "goodness-of-fit" statistics were comparable to the previous values summarized in Table 8-2, reflecting the stability of the four devised ANN models. For example, the R^2 value was evaluated to be 0.82, 0.84, 0.90, and 0.92 for the lumped, TDA = 250 ha, TDA = 100 ha, and TDA = 50 ha, respectively. Figure 8-11 presents a scatter plot of the measured versus the predicted TP concentrations above 50 μg/L for the cross-validation data sets before and after data swapping. It shows that all four models were successful in predicting TP concentrations above 50 μg/L and that the semi-distributed models were generally better than the lumped model. Models with TDA = 100 ha and TDA = 250 ha were better in predicting peak TP concentrations above 350 μg/L, and the model with the finest resolution (TDA = 50 ha) was the best overall.

When conducting a data-driven modelling approach, the parameters that are physically important might often turn out to be rather trivial in the actual modelling application. Moreover, sometimes the input parameters' importance changes significantly by changing the data set used in the model calibration, reflecting poorly tuned models. Thus, the relative importance of the model inputs with respect to their influence on the model output had to be assessed to ensure that the devised models are consistent with our conceptual understanding of the modelled system, and with each other. In this study, the "weights" method as described by Garson (1991) was used to identify the relative contribution of the inputs to the modelled daily TP concentration. The consistency in the importance of the model inputs was evident for all semi-distributed models (Figure 8-12). The relative importance of the summation of all model EVI and NDWI inputs (named VI in Figure 8-12) was close to 60% for all three semi-distributed models. The relative contribution of the daily streamflow inputs was close to 25%, ranking second after the VI inputs. The TP/Q hystereses inputs (named sin/cos in Figure 8-12) accounted for a 10% relative contribution, and the average air temperature ranked last in importance. The consistency among the three semi-distributed models reflects that the formulated models are robust representations of the modelled parameter. On the other hand, the lumped model performed differently, with streamflow being the most influential (45% relative contribution as opposed to 25% in the semi-distributed case) and with the VI inputs ranking second with a relative contribution of 25% (as opposed to 60% in the semi-distributed case). TP/Q hystereses and average air temperature relative importance were comparable in importance with their semi-distributed counterparts. Apparently, averaging the VI values over the entire watershed had a smoothing effect on the data that prevented the lumped model from extracting all possible relations between TP and both the EVI and NDWI. The lumped model had to rely more than the other models on streamflow to map TP (analogous to rainfall-runoff lumped models in hydrology). The prevalence of particulate phosphorus (PP) contribution of the TP time series and the high association between PP and Q made it easy for the lumped model to reasonably map TP in the absence of more detailed vegetation/soil information by giving more weight to the Q inputs.

As a final step, the spatial importance of the landscape vegetation in affecting water-phase TP concentration was studied. The sum of the relative contributions of the EVI and NDWI was calculated for each subwatershed. ArcGIS was used to group similar values and to display their variation on the map. An example output map for TDA of 50 ha is depicted in Figure 8-13. Such an output can be effectively used to guide forest management in a scenario-based analysis. It shows that, for the sake of preserving water quality, if we are to harvest and/or disturb a specified area of the watershed, we have to start with the yellow portion of the watershed, followed by the white, and finally the red. To move this analysis a step forward, a correlation between the currently used vegetation metrics and the RS VIs (the EVI and NDWI in this case) has to be formulated. Once the correlation is established, we can force disturbances in the landscape (by changing the values of the EVI and NDWI) and simulate the impact of the changes on the water quality.

8.7. Conclusions and Recommendations

This study developed four artificial neural network (ANN) models for daily total phosphorus (TP) predictions applied to the Willow watershed in Northern Alberta, Canada. Four years of data were used to calibrate and validate the models. The devised models were less reliant on ground based watershed-specific information and thus can be extrapolated for application to hydrologically similar watersheds. Four watershed subdivisions were incorporated to test the impact of watershed subdivision on the prediction accuracy of the ANN TP modelling. Eight measures of "goodness-of-fit" statistics including two with a penalty term for increasing model complexity were used to evaluate the developed models. Based on our modelling results, it was concluded that ANN modelling is well suited to model the daily TP concentration. The developed models utilized public domain remote sensing (RS) information available from the Moderate-resolution Imaging Spectroradiometer (MODIS) and typically available meteorological information as inputs for model building. The MODIS-derived Enhanced vegetation index (EVI) and the normalized difference water index (NDWI) were successful in representing vegetation dynamics in the devised models and, thus, in representing soil/vegetation phosphorus transport.

The results from the four models presented a good match of the measured and predicted TP profiles and an almost perfect material balance (a % relative bias of less than 3%). The statistical measures of the "goodness-of-fit" minimally favored the finest resolution semi-distributed model over other watershed subdivisions. In general, all the used measures of "goodness-of-fit" including the multivariate corrected Akaike's information criterion (AICc) and the Bayesian information criterion (BIC) were enhanced by increasing the degree of watershed subdivision. However, the slope of enhancement was minimal when the number of subwatersheds exceeded 7 subwatersheds.

Although the statistical model evaluation did favor the finest resolution in this case study, all model performance indicators were satisfactory for the four devised models. The differences in performance indicators were not significant, by any means, for any practical application. Therefore, it is concluded that the choice of the optimum watershed subdivision should depend upon the modelling objective. Lumped parameter modelling is easy to construct, relies on affordable landbase information, but cannot address questions related to the impact of different land use scenarios on water quality. Therefore, if the objective is to forecast real-time water quality or to assess the impact of climate change scenarios on water quality, lumped parameters modelling can be used for any practical purpose without jeopardizing prediction accuracy. On the other hand, if the objective is to quantify the impact of different land use activities, then the watershed must be divided into subwatersheds to make the model recognize the locations of disturbances and, thus, to be able to simulate the corresponding impacts on water quality. Based on our results, we conclude that only in this case is the added time, cost, and effort of preparing distributed landbase information and its subsequent data processing justifiable.

Although evidence and basic justifications for the obtained results were provided in this study, to strengthen our results even more, further investigations are needed to determine the influence of the impact of watershed discritization on watersheds of various sizes, and in different geomorphological and spatial settings, and with other water quality parameters.

8.8. References

Alberta Environment. 1999. Surface water quality guidelines for use in Alberta. Alberta Environment Pub. No.: T/483. ISBN: 0-7785-0897-8.

Alberta Sustainable Resource Development. 2006. Alberta Forest Management Planning Standard version 4.1.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R. 1998. Large area hydrologic modeling and assessment: Part 1. Model development. J. Am. Water Resour. Assoc. **34**: 73–89.

Beasley, D.B., Huggins, L.F., and Monke, E.J. 1980. ANSWERS: a model for watershed planning. Trans. ASAE **23**: 938–944.

Bedrick, E.J., and Tsai, C.L. 1994. Model selection for multi[illegible]ate regression in small samples. Biometrics 50: 226-231.

Bingner, R.L., Garbrecht, J., Arnold, J.G., and Srinivasan, R. 1997. Effect of watershed subdivision on simulation runoff and fine sediment yield. T ASAE. 40(5): 1329-1335.

Bingner, R.L., Theurer, F.D., Cronshey, R.G., and Darden, R.W. 2001. AGNPS 2001 Web Site. http://msa.ars.usda.gov/ms/oxford/nsl/AGNPS.html.

Bouraoui, F., and Dillaha, T.A. 1996. ANSWERS-2000: runoff and sediment transport model. J. Environ. Eng. **122**: 493– 502.

Boyle, D.P., Gupta, H.V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M. 2001. Toward improved streamflow forecasts: Value of semidistributed modelling, Water Resour. Res. **37**(11): 2749–2759.

Cheng, Y.B., Zarco-Tejada, P.J., Riano, D., Rueda, C.A., and Ustin, S.L. 2006. Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. Remote Sens. Environ. **105**(4): 354-366.

Chorus, I. (Ed). 2001. Cyanotoxins: occurrence, causes, consequences. Berlin: Springer Verlag.

Cook, D.J., Dickinson, W.T., and Rudra, R.P. 1985. GAMES—the Guelph Model for Evaluating the Effects of Agricultural Management Systems in Erosion and Sedimentation. User's Manual. Tech. Rep. 126-71. School of Engineering, Univ. of Guelph, Guelph, Ont.

Cooke, S.E., and Prepas, E.E. 1998. Stream phosphorus and nitrogen export from agricultural and forested watersheds on the Boreal Plain. Can. J. Fish. Aquat. Sci. **55**: 2292-2299.

Dawson, C.W., and Wilby, R.L. 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geog. **25**(1): 80-108.

Donnelly-Makowecki, L.M., and Moore, R.D. 1999. Hierarchical testing of three rainfall-runoff models in small forested catchments. J. Hydrol. **219**: 136–152.

Environment Canada. 2002. Digital archive of the Canadian climatological data (surface). Atmospheric Environment Service. Canadian Climate Centre, Data Management Division, Downsview, Ont.

Fairchild, J., and Leymarie, P. 1991. Drainage networks from grid digital elevation models. Water Resour. Res. **27**(5): 709–717.

FitzHugh, T.W., and Mackay, D.S. 2000. Impacts of input parameter spatial aggregation on an agricultural nonpoint source pollution model, J. Hydrol. 236: 35-53.

Gamal El-Din, A., and Smith, D.W. 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. **36**(5): 1115-1126.

Garson, G.D. 1991. Interpreting neural-network connection weights. AI Expert **4**: 47-51.

Haan, C.T. 1989. Parametric in hydrologic modelling. T ASAE. **32**(1): 137-146.

Hallengraeff, G.M. 1993. A review of harmful algal blooms and their apparent global increase. Phycologia **32**:79-99.

Hauhs, M., Neal, C., Hooper, R., and Christophersen, N. 1996. Summary of a workshop on ecosystem modelling: The end of an era?. Sci. Total Environ. **183**: 1–5.

Haykin, S. 1994. Neural networks: a comprehensive foundation, second ed. Macmillian CollegePublishing, Inc., New York.

Hoagland, P., Anderson, D.M., Kaoru, Y., and White, A.W. 2002. The economic effects of harmful algal blooms in the United States: estimates, assessment issues, and information needs. Estuaries **25**: 819–37.

Holmberg, M., Forsius, M., Starr, M., and Huttunen, M. 2006. An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams and impacts of climate change. Ecol. Modell. **195**(1-2): 51-60.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. **83**(1-2): 195-213.

Jha, M., Gassman, P.W., Secchi, S., Gu, R., and Arnold, J. 2004. Effect of watershed subdivision on SWAT flow, sediment, and nutrient predictions. J. Am. Water Resour. As. **40**(3): 811-825.

Johanson, R.C., Imhoff, J.C., Little, J.L., and Donigian, A.S. 1984. Hydrological Simulation Program-Fortran (HSPF) User's Manual. EPA-600/3-84-066. U.S. EPA, Athens, GA. 745 pp.

Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Wan, Z., Huete, A.R., Leeuwen, W.,V., Wolfe, R.E., Giglio, L., Muller, J., Lewis, P., and Barnsley, M.J. 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research. IEEE T. Geosci. Remote. **36**(4): 1228-1249.

Kalin, L., Govindaraju, R.S., and Hantush, M.M. 2003. Effect of geomorphological resolution on modelling of runoff hydrograph and sedimentograph over small watersheds. J. Hydrol. **276**: 89-111.

Landsberg, J.H. 2002. The effects of harmful algal blooms on aquatic organisms. Res. Fish. Sci **10**: 113–390.

Legates, D.R., and McCabe Jr., G.J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water. Resour. Res. **35**(1): 233-241.

Lek, S., Dimopoulos, I., and Fabre, A. 1996. Predicting phosphorus concentration and phosphorus load from watershed characteristics using backpropagation neural networks. Acta Oecol. **17**(1): 43-53.

Liu, Y.B., Corluy, J., Bahremand, A., De Smedt, F., Poorova, J., and Velcicka, L. 2006. Simulation of runoff and phosphorus transport in a Carpathian catchement, Slovakia. River Res. Applic. **22**: 1009-1022.

Maier, H.R., and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw. **15**: 101-124.

Mamillapalli, S. 1998. Effect of spatial variability on river basin stream flow modelling, Ph.D. Thesis, Purdue University, West Lafayette, IN.

Michaud, J., and Sorooshian, S. 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resour. Res. **30**(3): 593-605.

MODIS Land Science Team (MLST). 2004. MODLAND Product User Guides; MODIS Land Quality Assessment. Available from http://landweb.nascom.nasa.gov/cgibin/QA_WWW/newPage.cgifileName=modland_guide. [Accessed on 26 August 2004].

Munn, N., and Prepas, E.E. 1986. Seasonal dynamics of phosphorus partitioning and export in two streams in Alberta. Can. J. Fish. Aquat. Sci. **43**: 2464-2471.

Nash, J.E., and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, part 1: a discussion of principles. J. Hydrol.**10**: 282-290.

Norris J., and Hann, C.T. 1993. Impact of subdividing watersheds on estimated hydrographs. Appl. Eng. Agric. **9**(5):443-445.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006a. The application of artificial neural networks to flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands. Ecol. Model. **191**: 19-32.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2006b. Artificial neural networks and time series modelling of TP concentration in boreal streams: A comparative approach. J. Environ. Eng. Sci. **5**: S39-S52.

Nour, M.H., Smith, D.W., Gamal El-Din, M., and Prepas, E.E. 2007. Towards a Generic Artificial Neural Network Model for Dynamic Predictions of Daily Streamflow in Ungauged Watersheds: Introducing a new measure of hydrologic similarity. J. Environ. Eng. Sci. (Ch. 5. in review).

Refsgaard, J.C., and Knudsen, J., 1996. Operational validation and inter comparison of different types of hydrological models. Water Resour. Res. **32**: 2189–2202.

Schindler, D.W., Newbury, R.W., Beaty, K.G., Prokopowich, J., Ruszczynski, T., and Dalton, J.A. 1980. Effects of a windstorm and forest fire on chemical losses from forested watersheds and on the quality of receiving streams. Can. J. Fish. Aquat. Sci. **37**: 328-334.

Schwartz, F. 1978. Estimating the dimension of a model. Ann. Stat. 6: 461-464.

Sharpley, A.N., and Williams, J.R. (Eds.) 1990. EPIC—Erosion/Productivity Impact Calculator: I. Model Documentation, Tech. Bull., vol. 1768. U.S. Department of Agriculture.

Smith, D.W., Prepas, E.E., Putz, G., Burke, J.M., Meyer, W.L., and Whitson, I. 2003. The Forest Watershed and Riparian Disturbance study: a multi-discipline initiative to evaluate and manage watershed disturbance on the Boreal Plain of Canada. J. Environ. Eng. Sci. **2**: S1-S13.

Tripathi, M.P., Raghuwanshi, N.S., and Rao, G.P. 2006. Effect of watershed subdivision on simulation of water balance components. Hydrol. Process. **20**: 1137-1156.

Wilcox, B.P., Rawls, W.J., Brakensiek, D.L., and Wight, J.R. 1990. Predicting runoff from rangeland catchements: A comparison of two models. Water Res. Resear. **26**(10): 2401-2410.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J. and Rowe, C.M. 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. **90**: 8995-9005.

Zealand, C.M., Burn, D.H., and Simonovic, S.P. 1999. Short term stream flow forecasting using artificial neural networks. J. Hydrol. **214**: 32-48.

Zhang, W., and Montgomery, D.R. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulation. Water Resour.Resear. **30**(4): 1019-1028.

Zhang, Z., Koren, V., Smith, M., Reed, S., and Wang, D. 2004. Use of next generation weather radar data and basin disaggregation to improve continuous hydrograph simulations. J. Hydrol. Eng. **9**(2): 103–115.

Table 8-1. Summary table showing optimum ANN models' architecture and ANN internal parameters

	Model 1 (lumped)	Model 2 (TDA = 250 ha)	Model 3 (TDA = 100 ha)	Model 4 (TDA = 50 ha)
Data division (TS:SS:CVS)	3:1:1			
Scaling function	Linear, <-1,1>			
Optimum network (I-[H-H-H]-O)	9L-[4G-4GC-10LO]-T	17L-[4G-4GC-2LO]-T	21L-[3G-3GC-10LO]-T	29L-[4G-4GC-LO]-T
Training algorithm	Back-propagation			
Learning rate	0.1			
Momentum coefficient	0.1	0.1	0.1	0.02
Initial weights	Random [-0.3,0.3]			
Stopping criterion	Best test set (in terms of MSE)			

where: I and O denote input and output layers, respectively.; [H-H-H], represents a single hidden layer with different activation function; L, is the linear scaling function; G, GC, LO, and T are the Gaussian, Gaussian complement, logistic, and the hyperbolic tan activation functions, respectively; TS, SS, and CVS are training, testing, and cross-validation data sets, respectively; and < > denotes an open interval.

Table 8-2. Statistical measures of models' performance

	Model 1 (lumped)			Model 2 (TDA = 250 ha)			Model 3 (TDA = 100 ha)			Model 4 (TDA = 50 ha)		
	TS	SS	CVS	TS	SS	CVS	TS	SS	CVS	TS	SS	CVS
RB(%)	-2	-3	-2	3	-1	3	-1	-1	3	0	-2	0
RMSE	13.2	22.9	24.0	10.3	17.6	21.3	8.7	15.0	17.7	6.4	11.3	20.0
MAE	9.3	13.2	14.2	6.7	9.7	11.4	5.3	8.1	9.7	4.1	6.5	10.0
R^2	0.95	0.82	0.86	0.97	0.90	0.89	0.98	0.92	0.92	0.99	0.96	0.90
d_2	0.99	0.95	0.96	0.99	0.97	0.97	0.99	0.98	0.98	1.00	0.99	0.97
r^2	0.95	0.83	0.86	0.97	0.90	0.89	0.98	0.93	0.92	0.99	0.96	0.90
AIC$_c$		6.82			6.44			6.11			5.97	
BIC		5.82			5.45			5.13			5.00	

TS, training data set; SS, testing data set; CVS, cross-validation data set; RMSE and MAE are in μg/L. AIC$_c$ and BIC were calculated for all data patterns t compare between models.

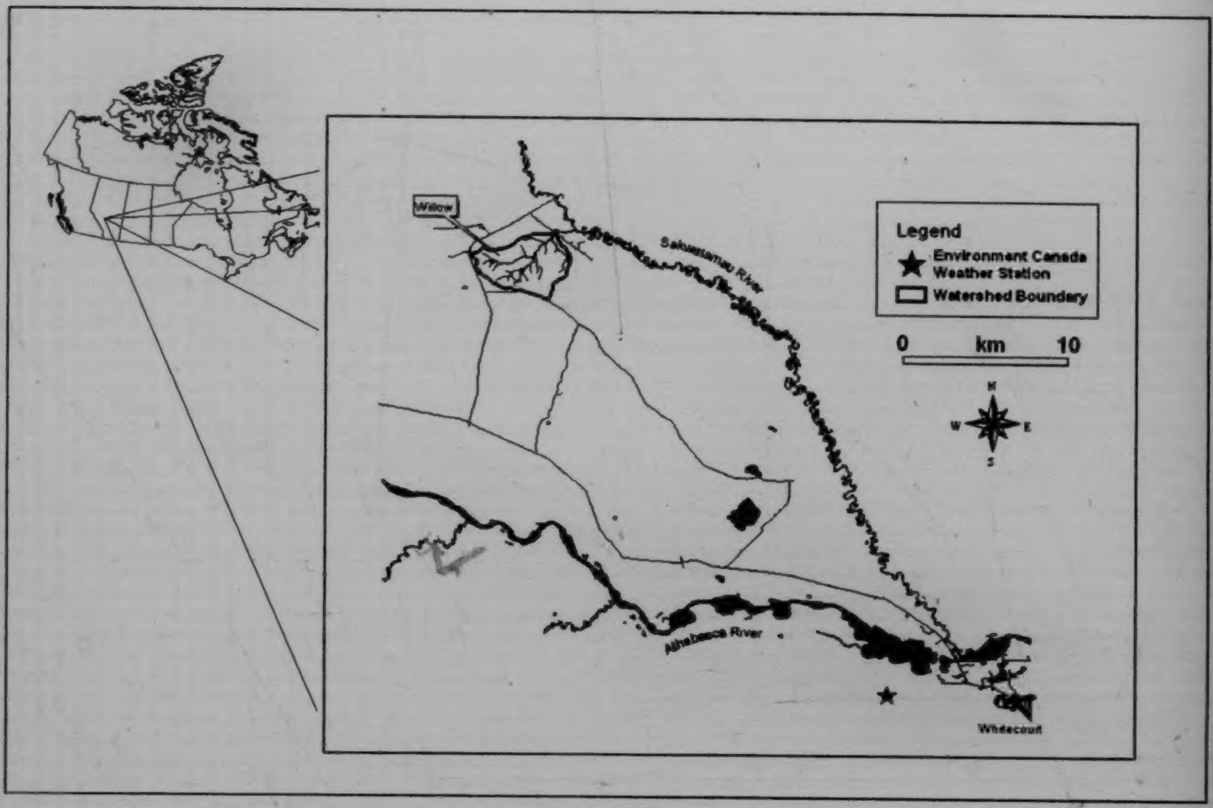

Willow
Sakwatamau River
Legend
Environment Canada Weather Station
Watershed Boundary
0 km 10
N
W
E
S
Athabasca River
Whitecourt

Figure 8-1. A schematic showing the studied watershed

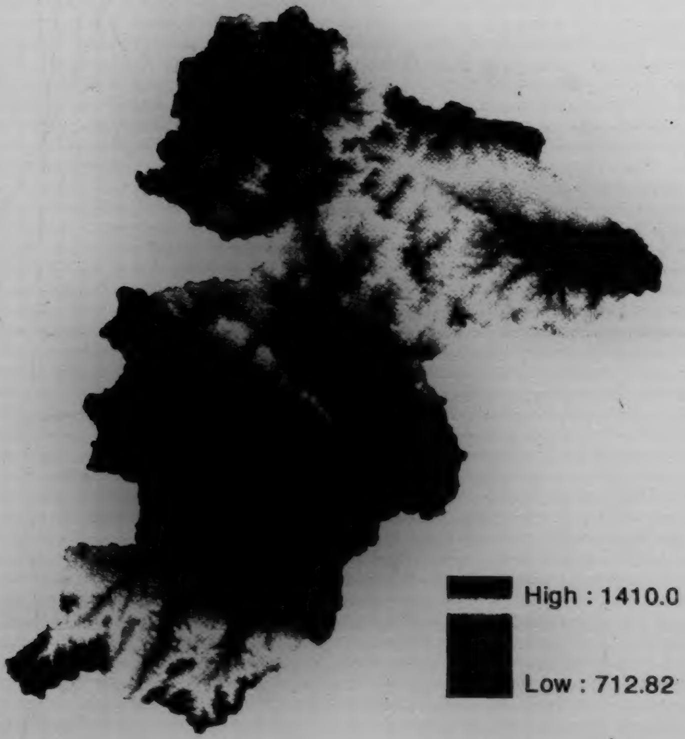

High : 1410.0
Low : 712.82

Figure 8-2. 30 m x 30 m resolution filled DEM of the FORWARD study area

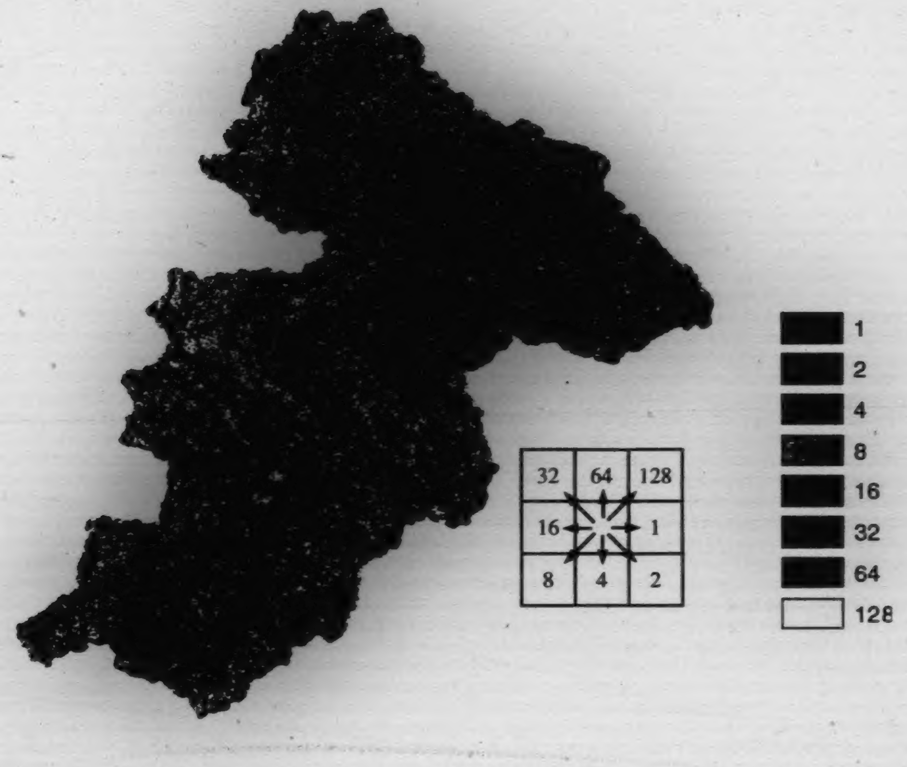

32
64
128
16
1
8
4
2
1
2
4
8
16
32
64
128

Figure 8-3. The FORWARD study area generated flow direction map using the D8 method

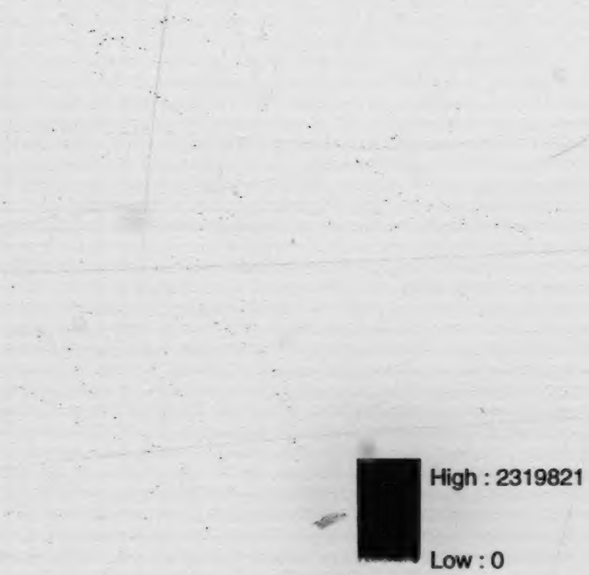

High : 2319821
Low : 0

Figure 8-4. Generated flow accumulation map for the FORWARD study area

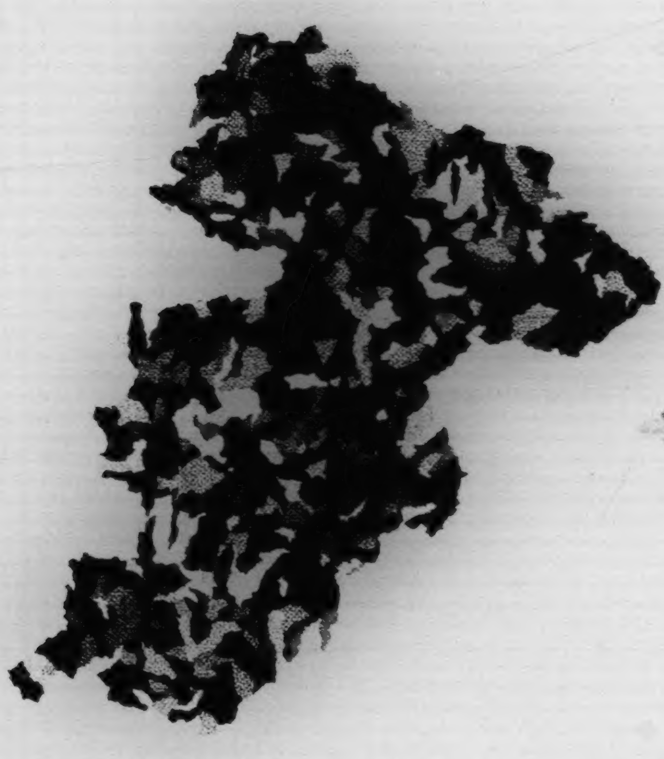


Figure 8-5. Generated watersheds for the FORWARD study area using a 250 ha TDA

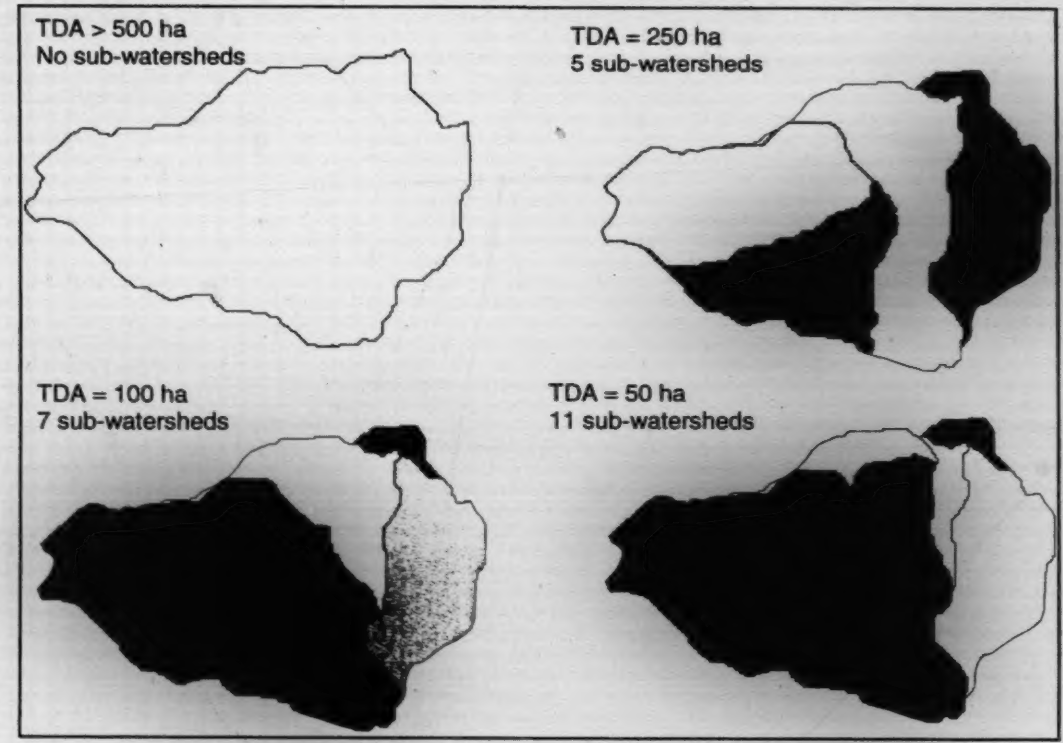

TDA > 500 ha
No sub-watersheds
TDA = 250 ha
5 sub-watersheds
TDA = 100 ha
7 sub-watersheds
TDA = 50 ha
11 sub-watersheds

Figure 8-6. The Willow watershed subdivisions used in this study

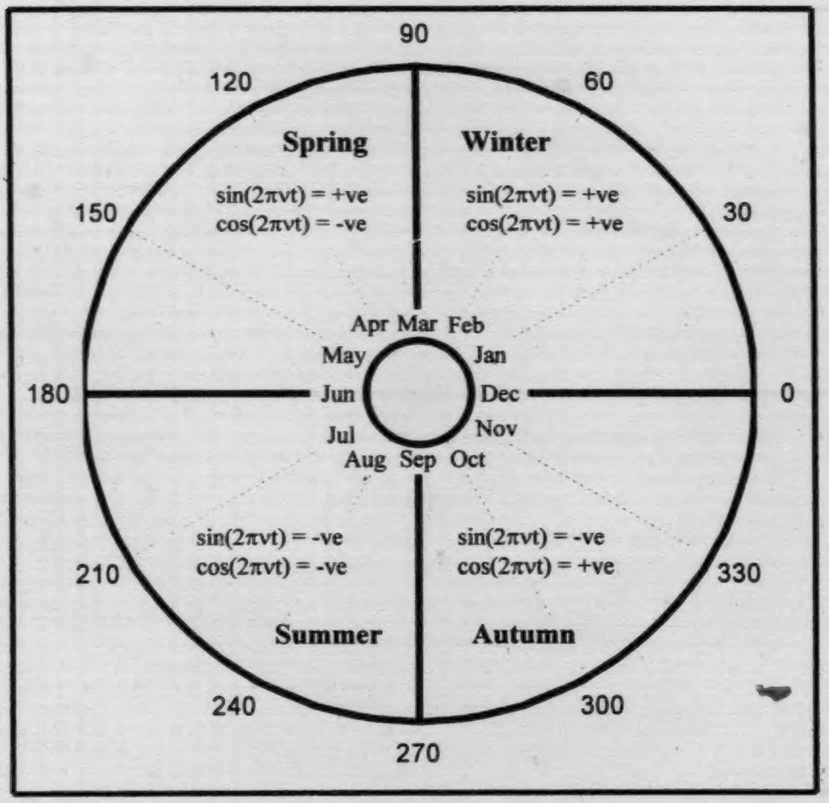

90
120
60
Spring
Winter
sin(2πvt) = +ve
cos(2πvt) = -ve
sin(2πvt) = +ve
cos(2πvt) = +ve
150
30
Apr
Mar
Feb
May
Jan
180
Jun
Dec
0
Jul
Nov
Aug
Sep
Oct
sin(2πvt) = -ve
cos(2πvt) = -ve
sin(2πvt) = -ve
cos(2πvt) = +ve
210
330
Summer
Autumn
240
300
270

Figure 8-7. Trigonometric quadrants describing the concept of feeding ANN TP models with seasonal variation and TP/Q hystereses loops

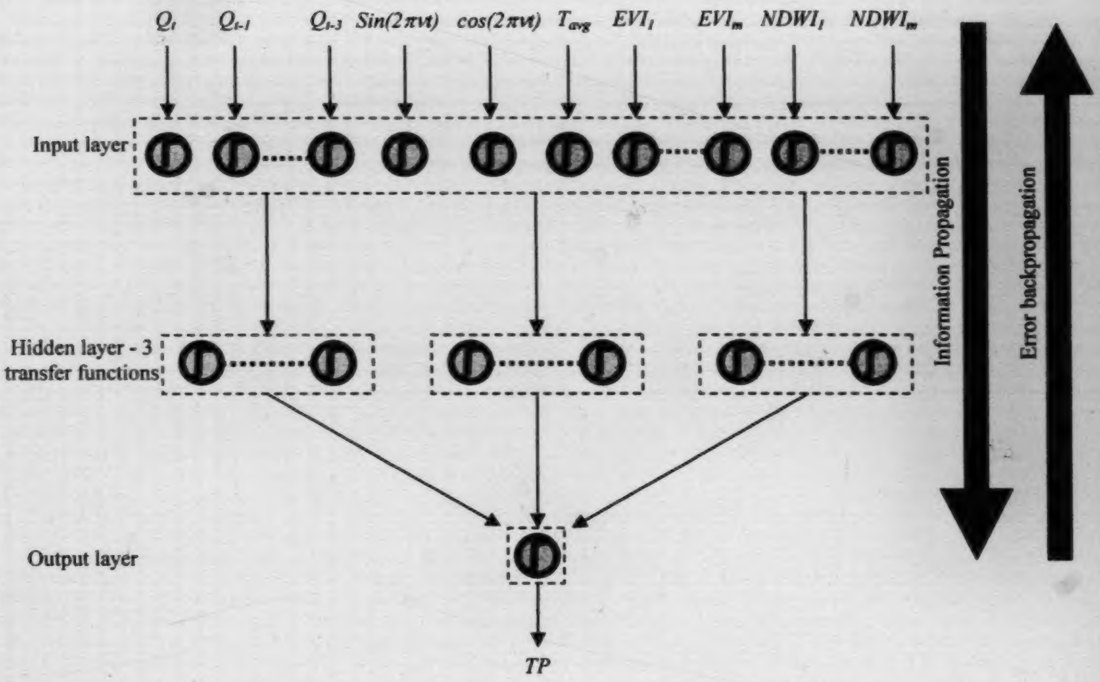

Q_t
Q_{t-1}
Q_{t-3}
$Sin(2\pi\nu t)$
$cos(2\pi\nu t)$
T_{avg}
EVI_l
EVI_m
$NDWI_l$
$NDWI_m$
Input layer
Hidden layer - 3 transfer functions
Output layer
TP
Information Propagation
Error backpropagation

Figure 8-8. ANN architecture used in this study

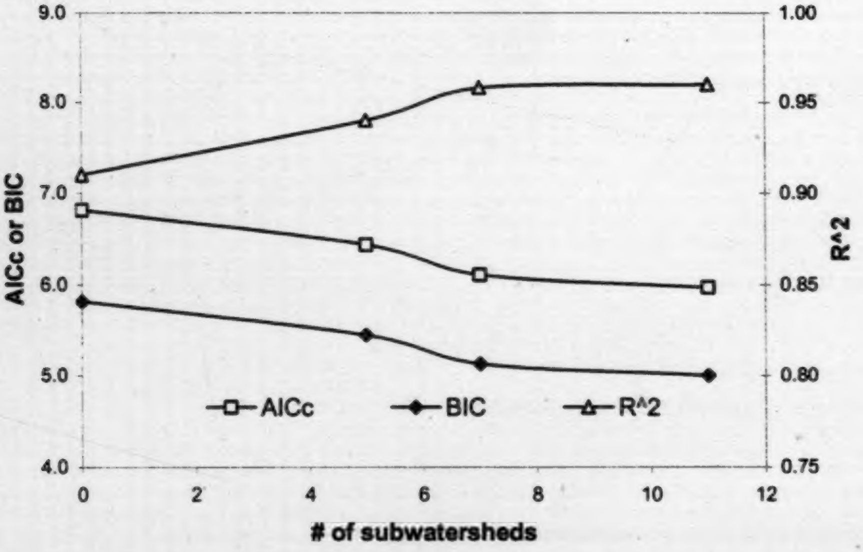


Figure 8-9. Impact of watershed discretizations on models' performance

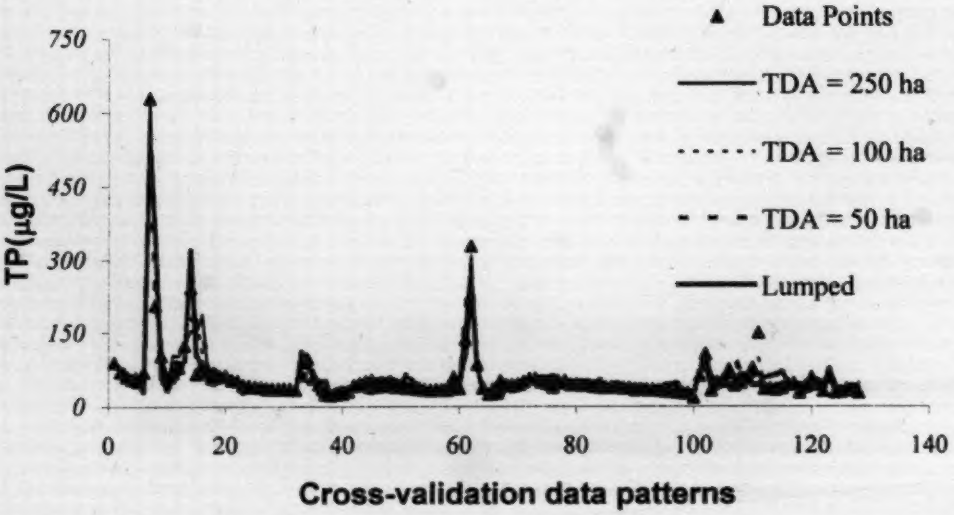


Figure 8-10. TP concentration profiles for all four models applied to the cross-validation data set

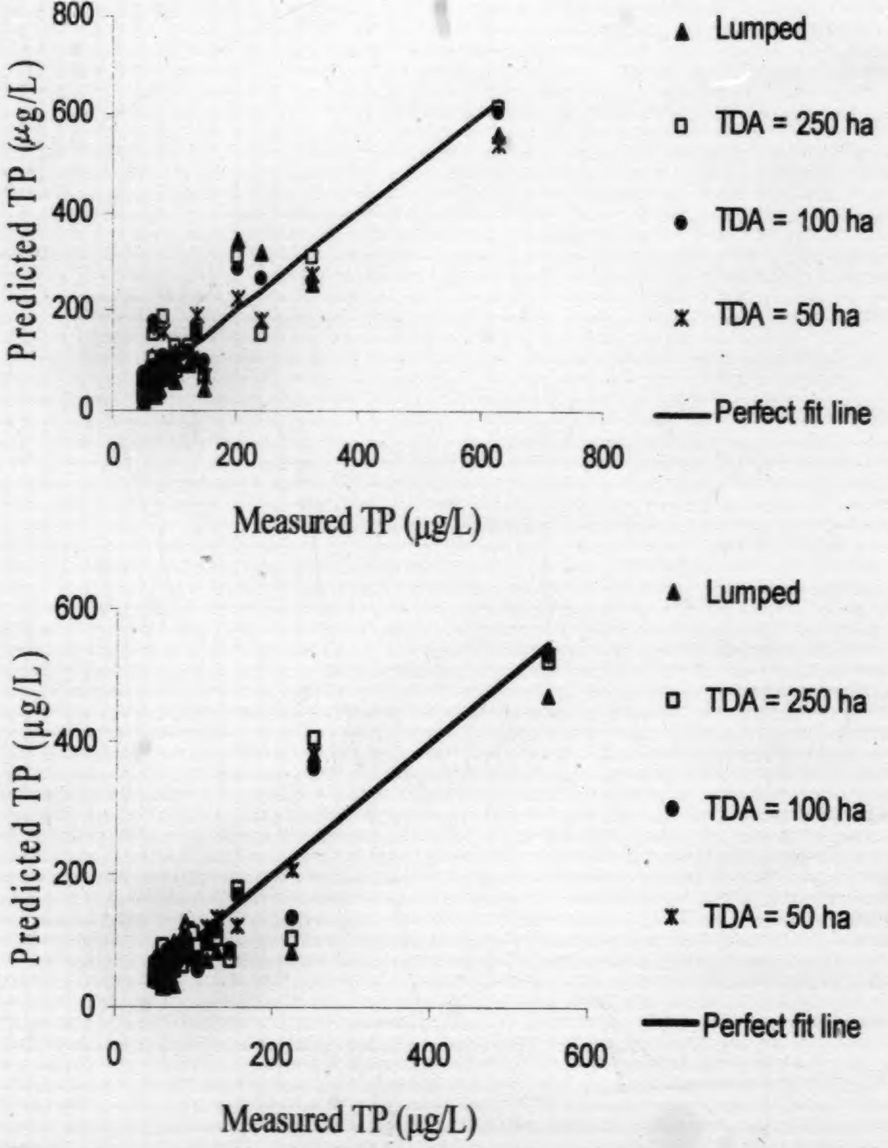


Figure 8-11. Predictions of TP concentration values > 50 μg/L in the cross-validation data set, SS used as the testing data set (top panel) and CVS used as the testing data set (bottom panel)

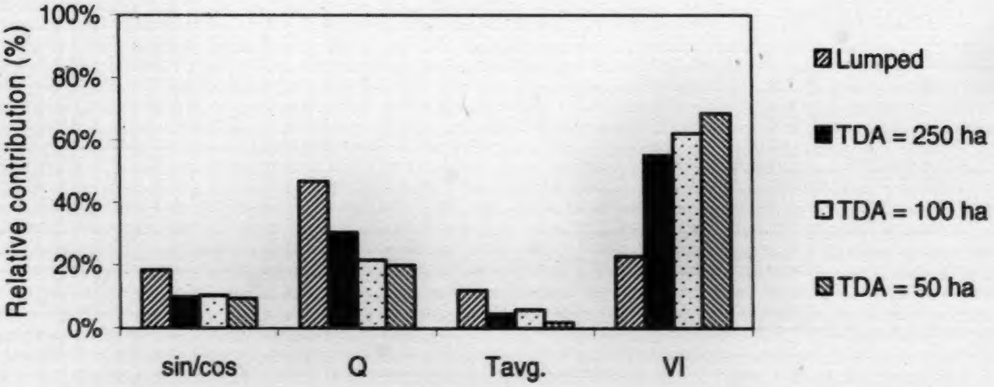


Figure 8-12. Relative contribution of inputs for all four models, sin/cos and VI are the sum of sine and cosine inputs and the sum of all RS vegetation indices inputs, respectively

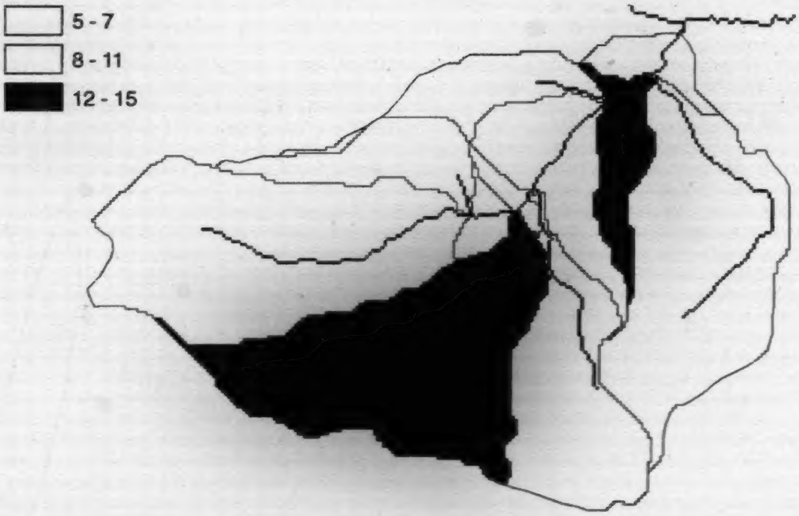


Figure 8-13. Spatial distribution of the relative contribution of the sum of EVI and NDWI for TDA of 50 ha

CHAPTER 9. GENERAL CONCLUSIONS AND RECOMMENDATIONS

9.1. Thesis Overview

In Canada and elsewhere, the prediction of daily streamflow (Q) and total phosphorus (TP) concentrations is important for evaluating downstream hydrologic impacts, simulating the impact of extreme floods and droughts, evaluating the impact of different climate change scenarios and, thus, for safeguarding against any expected adverse consequences. Providing the resources to gauge all watersheds of interest is practically impossible, therefore, most of the currently available models for watershed modelling are limited in practice because of the extensive requirement for landscape data (e.g. soils, vegetation, precipitation) needed for model calibration. Therefore, a class of models that can simulate the response of ungauged watersheds with reasonable accuracy is critical to provide the necessary information for responsive watershed management practices.

This thesis attempted to build a class of watershed models that are less reliant than currently available models on ground-based watershed specific measurements by using remotely sensed information instead. It investigated the possibility of developing a modelling approach capable of simulating streamflow and water quality in ungauged and unmonitored watersheds. The focus was on formulating Q and TP models that are only reliant on currently available meteorological information in Canada, as well as public-domain free-of-cost Moderate Resolution Imaging Spectroradiometer (MODIS) derived remote sensing (RS) information. A number of water quantity and quality models were devised and applied to a number of watersheds ranging in their basin area from 5 to 130 km^2. The protocols used for data pre-processing, model formulation, and model evaluation made the following original contributions: (1) This thesis represents the first effort to compare autoregressive moving average with exogenous input (ARMAX) modelling to artificial neural network (ANN) modelling for TP predictions. Earlier efforts had focused on comparing the two modelling approaches for Q predictions (Chapter 2); (2) the current study established step-by-step guidelines to ANN modelling of time-correlated variables that can account for data hystereses, as described in Chapter 2. The approach was tested further in Chapters 3, 5, 6, 7, and 8; (3) it proposed a feed-forward multi-layer perceptron (FF-MLP) modelling algorithm that relies only on low-cost, readily available meteorological data and careful time series manipulation prior to model building for Q predictions, and thus, is suitable for modelling streamflow in ungauged watersheds (Chapter 5); (4) in this thesis, a new remotely-sensed hydrologic similarity measure was proposed and was found to provide a successful indicator of basin similarity (Chapter 5); (5) the current research is the first to attempt to build a model that can rely on a dynamic suite of remotely sensed vegetation indices for predicting the water-phase TP concentration (Chapter 7); (6) it is also the first to address the impact of watershed subdivision on a water quality parameter using an ANN modelling algorithm (Chapter 8).

In the preceding eight chapters, thorough discussions of the pertinent work were presented and the specific conclusions of each chapter were drawn. This chapter presents briefly the general conclusions of this study, as well as the recommendations for possible future work.

9.2. Conclusions

(1) In this study, ARMAX modelling was compared to ANN modelling for TP concentration predictions. Both ARMAX and ANN predicted TP concentration with reasonable accuracy. Based on four measures of goodness-of-fit statistics, including two with a penalty term for increasing model complexity, and by examining the predicted versus the measured TP concentration profiles for the two modelling approaches, it was evident that ANN outperformed ARMAX. The inherent nonlinearity of the modelled process is likely to favor the ANN approach. Whereas ANN was capable of modelling gaped data efficiently, the ARMAX approach required equally spaced data values; therefore values had to be estimated for times when measurements were not available for the ARMAX model development.

(2) A multi-slab ANN was designed and utilized in this study in which a three-slab hidden layer with three different activation functions was used to reflect the distinct system behavior with respect to base flow, snow melt, and rain events. The approach was found useful throughout all devised models in this thesis and, thus, can be proposed as an improved ANN architecture for modelling streamflow and water quality parameters in northern latitudes.

(3) The hystereses in TP/Q and in Q/R behaviors were adequately addressed in this thesis by first conducting spectral analysis on the data and then by introducing two additional hystereses-specific inputs into the ANN model development.

(4) This study also proposed systematic guidelines for modelling time-correlated variables that can account for data hystereses. These guidelines can potentially be

applied for modelling other water quality parameters including diffuse pollutants associated with the melting of contaminated snow packs.

(5) The rainfall time series is the most important input in hydrologic water quantity and quality models. Thus, securing the highest-possible rainfall data quality has always been a priority in order to ensure the development of reliable water quality and streamflow predictions models. Practitioners often get confused in the realm of kriging methods and other interpolation schemes available. This thesis examined five geostatistical interpolation techniques; simple kriging (SK), ordinary kriging (OK), multi-Gaussian kriging (MGK), log-normal kriging (LNK), and kriging with an external drift (KED), in addition to the more traditional inverse distance weight (IDW) interpolation technique for the estimation of daily rainfall in a 250 m x 250 m grid over a 750 Km^2 area in the Canadian Boreal forest. The results indicated that multivariate kriging did not enhance the daily rainfall estimation skill. This can be explained by the relatively mild slope of the study area. SK, OK, LNK, and IDW were proven to be comparable algorithms with respect to prediction performance; however, OK produced slightly better results in terms of Pearson's correlation coefficient (r). IDW outperformed OK in terms of RMSE but came next to it in performance with regards to *r*. MGC was the worst univariate estimator, likely due to the high percentage of data spikes. Although data despiking was carefully done, Gaussian back-transfer was disrupted due to the high proportion of data spikes.

(6) The strength of OK, as compared to IDW, was in the ability to estimate a measure of prediction uncertainty by evaluating the kriging variance. However, for our case study, because a prediction error estimate is not required, IDW was used for

estimating daily rainfall from 15 Environment Canada and fire towers weather stations.

(7) Sequential Gaussian Simulations (SGS) was then implemented to produce 100 equiprobable maps of rainfall and the probability of exceeding nominal thresholds of rainfall over the study domain was calculated. Such information can be used as inputs to hydrologic and water quality models in order to address the uncertainty in the modelled parameter in response to the uncertainty in rainfall information.

(8) A multi-objective approach for selecting future rain gauge sites; that is based on overlaying the map of the kriging variance, the digital elevation model (DEM), and the land use/land cover and road networking maps in a GIS framework to identify the areas of commonly favorable features; was proposed to identify potential future sampling locations. The approach was applied to the FORWARD study area and favorable regions for the installation of further weather stations were identified.

(9) Most of the reviewed streamflow neural network models were either recurrent network based or feed-forward multi-layer perceptron (FF-MLP), requiring the past flow values for lead-time prediction. These models cannot be used in modelling ungauged watersheds when such information is missing. The current study proposed a FF-MLP algorithm using low-cost, readily available meteorological data and careful time series manipulation prior to model building. The proposed algorithm used IDW interpolation for better rainfall representation. The temperature index snowmelt approach was used to account for the snowmelt. Cross correlation analysis was used to identify the time-lagged inputs, and spectral analysis was used to feed the model with information representing the Q/R hystereses loops. The

algorithm was applied to four watersheds (5 to 130 Km^2 of basin area) in the Canadian Boreal Plain. All models managed to simulate streamflow fairly well at all data ranges. Six measures of the "goodness-of-fit" were used to assure model accuracy. To demonstrate the approach's applicability to modelling ungauged watersheds, the calibrated models were applied to a smaller watershed, the Mosquito watershed. In addition, the model initially developed for the 1A watershed (5.1 km^2) was used in a predictive mode to simulate three years of streamflow for the Cassidy watershed (5.9 km^2). The initial results from these applications are very promising.

(10) A new hydrologic similarity index (SWMIR_SIi,j) that makes use of public domain remote sensing information was proposed and was found to be significantly correlated to model performance.

(11) In this study, we attempted to build a model that relies on a time series of remotely sensed vegetation indices for predicting the dynamics of water-phase TP concentration. Because of the different degrees of sensitivity of each vegetation index (VI) to ecosystem disruption and plant health (as a responsive indicator of soil nutrient availability and other growth limiting factors), we examined the possibility of using seven VIs; enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), greenness fraction vegetation index (GFVI), simple ratio vegetation index (SRVI), normalized difference water index (NDWI), in addition to two indices proposed in this study (SRVIm and GFVIm) to provide sufficient landscape information for water-phase TP modeling. In addition, we conducted a sensitivity analyses to quantify the uncertainty in model predictions in response to

the uncertainty in these VIs. A protocol for constructing parsimonious ANN TP models and for linking RS VIs to ANN was developed. EVI-based models were superior in terms of the model prediction accuracy, model robustness and stability. The sensitivity analysis confirmed the robustness of the model to possible errors in EVI values and its sensitivity to vegetation disruption in response to watershed disturbance practices. Building TP models that rely on both EVI, as a representative to vegetation chlorophyll content, and NDWI, representing vegetation water content, proved to be a superior modelling alternative. The improved model responses can be attributed to the incorporation of the MODIS derived VIs in the modelling exercise and their success in reflecting the vegetation dynamics and the dynamics of the soil/vegetation phosphorus transport.

(12) The impact of watershed subdivision on modelling TP concentrations was also studied in this thesis. Four watershed subdivisions were incorporated to test the impact of watershed subdivision on the prediction accuracy of the ANN TP modelling. Eight measures of the "goodness-of-fit" statistics including two that have a penalty term for increasing model complexity were used to evaluate the developed models. The results of the four models presented a good match of the measured and the predicted TP profiles and an almost perfect material balance (a percentage relative bias of less than 3% was achieved). The statistical measures of the "goodness-of-fit" minimally favored the finest resolution semi-distributed model to the other watershed subdivisions. In general, all the used measures of "goodness-of-fit" including the multivariate corrected Akaike's information criterion (AICc) and the Bayesian information criterion (BIC) were enhanced by increasing the degree of

watershed subdivision. However, the slope of enhancement was minimal when the number of subwatersheds exceeded 7 subwatersheds. Although the statistical model evaluation did favor the finest resolution in this case study, all model performance indicators were satisfactory for the four devised models. The differences in performance indicators were not significant for any practical application. Therefore, it was concluded that the choice of the optimum watershed subdivision should depend upon the modelling objective. Lumped parameter modelling is easy to construct, relies on affordable landbase information, but cannot address questions related to the impact of different land use scenarios on water quality. Therefore, if the objective is to forecast real-time water quality or to assess the impact of climate change scenarios on water quality, lumped parameters modelling can be used for any practical purpose without jeopardizing prediction accuracy. On the other hand, if the objective is to quantify the impact of different land use activities, then the watershed must be divided into subwatersheds to make the model recognize the locations of disturbances and, thus, to be able to simulate the corresponding impacts on water quality. Based on our results, we conclude that only in this case is the added time, cost, and effort of preparing distributed landbase information and its subsequent data processing justifiable.

(13) Models like the ones developed in this study that use commonly available inputs, yet are reasonably accurate, provide a useful tool for modeling ungauged watersheds. The concepts presented in this study can easily be extrapolated to other similar watersheds permitting flow and water quality predictions in response to climate change and landscape management practices. It can also offer a hydrologic

link to the development of multi-objective forest management plans. Moreover, probabilities of exceeding a certain threshold of a contaminant as a result of a specific rare storm event (like the occurrence of 1/30 years storm event) can also be calculated by such models.

9.3. Recommendations for Future Work

Like all techniques, ours has room for improvements. Thus, a number of recommendations can be made for future research in this topic:

(1) The division of the hidden layer into three slabs, each using a different activation function, was found to be the best ANN architecture in all the devised models. It is believed that each slab can manipulate one of the three distinct processes that control the system behavior (base flow, snowmelt, and rain events). However, this thesis did not attempt to examine this hypothesis. Future research examining the relationship between the different types of streamflow hydrographs and neural networks' architectures can shed more light on to the abovementioned hypothesis.

(2) The results of TP concentration predictions for a watershed with a relatively high percentage of wetland aerial coverage (1A Creek) suggested that wetland-specific inputs are needed for better ANN model predictions. Therefore, more information about the dynamics of phosphorus export from at least some wetland types is required for better representation of wetland characteristics in the development of future TP ANN models.

(3) This thesis proposed a new remotely sensed measure of hydrologic similarity. While evidence and basic justifications for the obtained results were provided, to strengthen our results even more, additional applications to other watersheds is required to investigate the advantages and the limitations of the proposed hydrologic similarity index.

(4) The proposed hydrologic similarity index can potentially provide a good measure of the potential of streamflow model transferability from one watershed to the other. Future investigations should test the applicability of using the proposed index when modelling water quality parameters as well.

(5) A multi-objective approach for the selection of future weather stations' locations was proposed in this study. Favorable regions for the installation of further weather stations were then identified. Other measurements of interest (like those of air temperature and solar radiation) typically take place in each weather station. In future research, therefore, the same procedure should be conducted for the other parameters of interest as well. Upon completing the study of all the parameters of interest, an optimization algorithm targeting the minimization of a cost function, which reduces the cost and the kriging variance by choosing the optimal locations of the new stations, can be conducted for the areas that are commonly favorable for all the parameters of interest. Such an optimization could be a topic of a future research investigation.

(6) A study designed to investigate the impact of developing individual season-specific interpolation models and to compare their prediction performance to that of the general model devised in this study can be beneficial.

(7) The present study represents as a leading effort in incorporating low-cost time-variant information in constructing a predictive tool for water-phase TP concentration. It can potentially provide a valuable tool for simulating the impact of different watershed harvesting activities on water quality parameters. Further studies are required to confirm the applicability of the approach for different landscape dominated watersheds (e.g. urbanized, harvested, and burnt watersheds). In particular, low density vegetated watersheds should be examined as NDVI-based models may perform better than other models.

(8) To move this analysis a step forward, a correlation between the currently used vegetation metrics (e.g., timber supply, average height, average age, average diameter at breast height) and the RS VIs (like EVI and NDWI) has to be formulated. Once the correlation is established, we can force a disturbance in the landscape (by changing the values of such VIs) and then simulate the consequent impact of these changes on water quality.

(9) Although evidence and basic justifications for the obtained results were provided in this study, to strengthen the results even more, further investigations are needed to determine the influence of the impact of watershed discritization on watersheds of various sizes, and in different geomorphological and spatial settings, and with other water quality parameters.

(10) The possibility of linking the devised ANN models with another physically-based modelling approach that could ensure physically plausible model predictions by creating boundaries on ANN model parameters is another area for future research.

(11) The present study has shown the potential uses and benefits of some of the currently available satellite images. However, many new satellites valuable to watershed modelling either have been launched recently, or are to be launched in the near future. Improved estimation of vegetation characteristics, precipitation, soil moisture, inundated areas and snow characteristics seems to be an immediate outcome of this development. Future research should test the applicability of using such information in water quality modelling, and should address the future cost savings from incorporating RS information into modelling.

(12) Cross-correlation analysis was used to account for the time series effect of the model inputs. Despite the apparent success of the approach, as demonstrated in this study and in others, other nonlinear dependence techniques like the mutual information criterion (MIC) may prove more useful in accounting for the time-correlated variables in the model development. Future work should investigate whether the added effort in computing the MIC is justifiable in terms of the enhancement of the model's prediction capability.